National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

UNIVERSITY OF ALBERTA

PARAMETER ESTIMATION IN NMR USING LPSVD

BY

© EDMUND KA-YUEN HO

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF APPLIED SCIENCES IN MEDICINE

EDMONTON, ALBERTA
SPRING 1991

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-66581-5

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: EDMUND KA-YUEN HO

TITLE OF THESIS: PARAMETER ESTIMATION IN NMR USING LPSVD

DEGREE: MASTER OF SCIENCE

YEAR THIS DEGREE GRANTED: SPRING 1991

Permission is hereby granted to the UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend such copies of this thesis for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

PERMANENT ADDRESS:

118 Clansman Blvd.,
Willowdale, Ontario
M1W 2V5.

Date: January 15, 1991.

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled PARAMETER ESTIMATION IN NMR USING LPSVD submitted by EDMUND KA-YUEN HO in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE.

P. S. Allen
....
Dr. P.S. Allen (Supervisor)

....
Dr. R.E. Snyder (Supervisor)

Z J Koles
....
Dr. Z.J. Koles

....
Dr. H.R. Hooper

Date: January 15, 1991

Praise the LORD.

Praise God in his sanctuary;
praise him in his mighty heavens.

Praise him for his acts of power;
praise him for his surpassing greatness.

Praise him with the sounding of the trumpet,
praise him with the harp and lyre,
praise him with tambourine and dancing,
praise him with the strings and flute,
praise him with the clash of cymbals,
praise him with resounding cymbals.

Let everything that has breath praise the LORD.

Praise the LORD.

Psalm 150

This thesis is dedicated to my wife Ivy, and our parents

ABSTRACT

Three methods were compared for estimating the parameters of a nuclear magnetic resonance free induction decay (FID) signal composed of an unknown number of complex exponentially-damped sinusoidal signals in white Gaussian noise: the conventional discrete Fourier transformation (DFT) method, the Tufts-Kumaresan (TK) algorithm, also called the linear prediction singular value decomposition (LPSVD) algorithm, and a proposed method which uses estimates from the LPSVD algorithm as initial estimates for a Levenberg-Marquardt nonlinear least squares algorithm. Comparisons were made using simulated FID signals, each composed of two or three sinusoidal signals to which had been added white Gaussian noise of varying intensity.

The biases (mean of estimated less true value), variances and mean-square errors (MSE's; variance plus square of the bias) of the parameters estimated by the three methods were compared to one another, to values predicted by first order perturbation analysis of the LPSVD algorithm, and to the Cramer-Rao lower limits. Our results showed the LPSVD algorithm to yield variances of the estimated parameters consistent with those predicted by perturbation analysis, but significant parameter biases, in contrast to zero bias for all parameters predicted by perturbation analysis. In general, the proposed method, using 150 data points and an LPSVD tentative model order of approximately 75, out performed the 150-point LPSVD algorithm, which in turn out performed a 1024-point DFT method. Both the LPSVD method and the proposed method were able to estimate the signal parameters more accurately with significantly fewer data points than was the DFT method. In all simulations tested, the proposed method yielded variances and MSE's equal, within statistical uncertainty, to the Cramer-Rao lower limits.

ACKNOWLEDGEMENTS

Praise be to the God and Father of our Lord Jesus Christ, the Father of compassion and the God of comfort, who comforts us in all our troubles, . . . It is His wonderful grace that made this thesis a reality.

I would like to thank my supervisors Dr. Peter S. Allen and Dr. Richard E. Snyder for their continuous support and encouragement throughout my graduate education. They provided numerous constructive criticisms and expertly guided the many revisions.

I am also grateful to Professors Zole J. Koles, Rick Hooper and Larry Filipow who served as members on my thesis defense committee. They give me new insight to the thesis material.

I also like to thank Dr. Carey Gibson, Mr. Grey Shaw and Dr. Eden Wu for their encouragement and many helpful discussions which enabled me to see digital signal processing and statistics from different perspective. I am very happy to acknowledge my debt to them. I must also thank all my friends in the *in-vivo* NMR Unit who helped and encouraged me in doing the research.

I am also indebted to Mr. Simon Ho and my brothers and sisters in Christ, especially Billy Lui, Stephen Tam, Stephen Tung, Simon Wan, Titus Yu, and their families, and others in the Luke fellowship of the Edmonton Chinese Alliance Church, for their continuous prayer supports.

My appreciation for my wife Ivy for patiently supporting me in the last few years cannot be overstated. I was lucky enough to have a loving home while studying.

TABLE OF CONTENTS

CHAPTER PAGE

1 Introduction 1
  1.1 Definition of the problem 1
  1.2 Organization of the thesis 2
  1.3 History of the problem 2
  1.4 Some current methods 3
    1.4.1 Maximum likelihood estimation (MLE) 4
    1.4.2 Damping rate and frequency estimation by spectral estimation 9
      1.4.2.1 Prony's method and some close relatives 10
    1.4.3 Principal component estimation 13
    1.4.4 Noise subspace estimation 15
      1.4.4.1 Pisarenko harmonic decomposition (PHD) 16
  1.5 Objectives of the thesis 18

2 The basis of NMR 20
  2.1 Introduction 20
  2.2 Simple resonance theory 20
  2.3 Interaction between magnetization and magnetic field 25
    2.3.1 Motion in a static magnetic field : Larmor precession 26
    2.3.2 The influence of a rotating field : resonance 27
  2.4 The chemical shift 31
  2.5 Relaxation phenomena 32
    2.5.1 Longitudinal relaxation 33
    2.5.2 Transverse relaxation 38
  2.6 Bloch equations in the rotating frame 39
    2.6.1 Ideal pulse experiment 41
    2.6.2 The effect of finite RF pulse amplitude 43
  2.7 Detection of NMR signals 48
  2.8 Conventional signal processing in pulsed NMR spectroscopy 53
    2.8.1 Fourier transformation 53
    2.8.2 Zero filling 56

2.8.3 Baseline correction 57
2.8.4 Phase correction 58
2.8.5 Apodization 59
2.8.6 Resolution enhancement 60

3 Pseudo-inversion and singular value decomposition 62
3.1 Introduction 62
3.2 Pseudo-inversion and minimum 2-norm least squares solution 64
3.3 Singular value decomposition (SVD) 69
3.4 SVD and rank determination 73

4 Two algorithms 82
4.1 Introduction 82
4.2 LPSVD method 82
4.2.1 Noiseless case 84
4.2.2 Perturbation analysis 90
4.3 A proposed method 97

5 Simulation results 99
5.1 Introduction 99
5.2 Number of tentative model order and data points for the LPSVD method and the proposed method 102
5.3 Indistinguishable signal and noise singular values 123
5.4 Effect of noise (i.e., spectral sensitivity) 141
5.5 Effect of peak separation (i.e., spectral resolution) 155
5.6 Effect of changing a peak's damping rate (i.e., T_2 dependence) 168
5.7 Effect of changing a peak's amplitude 181
5.8 Effect of adding a well separated peak 195
5.9 Effect of variable baseline (i.e., broad underlying signal) 213

6 Conclusions 226

References 233

Appendix 1 Theoretical autocorrelation matrix of M complex exponentially damped sinusoids.................................239

Appendix 2 Proof of generalized Pisarenko property for damped sinusoids...241

Appendix 3 Definition of NMR spectroscopy frequency in unit "ppm" 242

Appendix 4 First order approximation of the discrete Fourier transform of complex damped sinusoid 243

Appendix 5 Amplitude response of window function in decibel scale 245

Appendix 6 Proof of Theorem 3.2.1...246

Appendix 7 Proof of Theorem 3.2.2...247

Appendix 8 Proof of Theorem 3.2.4...248

Appendix 9 Proof of Theorem 4.2.1.1 250

Appendix 10 Proof of Theorem 4.2.1.2 253

Appendix 11 Proof of Theorem 4.2.1.3 256

Appendix 12 Proof of Theorem 4.2.2.1 261

Appendix 13 Proof of Equation (4.2.2.26)......................................263

Appendix 14 Proof of property 2 in chapter 4.................................267

Appendix 15 Derivation of Cramer-Rao (CR) bounds for the parameters of M multiple complex exponentially damped signals in white Gaussian noise..270

Appendix 16 Derivation of the magnitude square area of an exponentially decaying signal peak...273

Appendix 17 Derivation of the area under the real part of the spectrum for an exponential decaying signal 274

Appendix 18 Linewidth of an exponentially damped signal 276

Appendix 19 Derivation of time domain signal-to-noise ratio for an exponentially decaying signal embedded in complex white Gaussian noise 278

Appendix 20 Statistical relationship of zero mean white Gaussian noise in time domain and frequency domain 280

Appendix 21 Levenberg-Marquardt method of nonlinear least squares problem 282

LIST OF TABLES

Table 2.4.1 Chemical shifts observed in high-resolution and in in-vivo spectroscopy. 32

Table 2.6.2.1 Approximate spectral width and B_1 field strength in in-vivo spectroscopy, assuming $B_0 = 2.35$ T and B_1 is in the range of 10^{-5} T to 10^{-3} T. 46

Table 3.4.1 Demonstration of the effect of noise on the singular values of $\hat{A}$ given by Equation (3.1.3). 74

Table 3.4.2 Influence of near rank deficiency and noise on the singular values of $\hat{A}$. 78

Table 5.1 Parameters that will be compared by different methods in this chapter. 100

Table 5.2.1.1 Optimal range of the tentative model L for the LPSVD algorithm with N = 50. 109

Table 5.2.2.1 Optimal range of the tentative model L for the LPSVD algorithm with N = 100. 110

LIST OF FIGURES

Figure 1.4.1 $|\beta| = |e_{c1}^{H} e_{c2}|$ verses normalized frequency separation of two peaks for three different sets of $\text{œ}_1 + \text{œ}_2$ 8

Figure 2.2.1 Angular momentum vectors. 22

Figure 2.2.2 Magnetic energy levels for nuclei of spin $\frac{1}{2}$ and 1. 23

Figure 2.2.3 The splitting of magnetic energy levels of protons, expressed as resonance frequency υ_0, as a function of magnetic field strength. 24

Figure 2.3.1.1 Precession of a classical angular momentum **J**, having a proportional magnetization $\mathbf{M} = \gamma \mathbf{J}$, about a static magnetic field $\mathbf{B}_0$. . 27

Figure 2.3.2.1 Fixed laboratory coordinate system Oxyz, and the rotating system about Oz with the angular velocity ω_{rf}, Ox'y'z'. 29

Figure 2.3.2.2 In the rotating frame Ox'y'z', the effective field $\mathbf{B}_{\mathbf{eff}}$ has a fixed direction, about which the magnetic moment **M**(t) rotates with a constant angular velocity about $\mathbf{B}_{\mathbf{eff}}$.. 30

Figure 2.3.2.3 (a) Nutation of **M** away from its equilibrium position onto the rotating frame produced by a 90° RF pulse. (b) Continued free precession of **M** in the laboratory frame after the termination of the RF pulse. 31

Figure 2.5.1.1 Motional frequency spectrum. 37

Figure 2.5.2.1 Dephasing of spins in the rotating frame due to differences in the local magnetic field felt by the various spins. 39

Figure 2.6.2.1 Tilted effective field in the rotating frame. 44

Figure 2.6.2.2 Effect of frequency offset $\Omega/(\gamma B_1)$ on the signal phase and effective pulse rotation angle β_{eff} for a single-pulse experiment.......... 47

Figure 2.6.2.3 Effect of frequency offset $\Omega/(\gamma B_1)$ on transverse magnetization $M^r_{xy}(0_+)$ for a single-pulse experiment. 47

Figure 2.7.1.1 A simplified block diagram of a pulsed NMR spectrometer receiver. 51

Figure 2.8.1.1 Illustration of a Fourier transform pair. 55

Figure 2.8.3.1 Illustration of one method of baseline correction. 58

Figure 3.4.1 The range of simulated spectra used to demonstrate the effect of noise on the singular values. 75

Figure 3.4.2 Graphical presentation of the results in Table 3.4.1. 76

Figure 3.4.3 The range of simulated spectra used to demonstrate the effect of near rank deficiency and noise on the singular values. 79

Figure 3.4.4 Graphical presentation of the results in Table 3.4.2. 80

Figure 4.2.1.1 Demonstration of the zeros of G(z) obtained from the LPSVD algorithm 89

Figure 5.2.1.1 A representative 1024-point DFT spectrum for Examples 5.2.1 to 5.2.3. 103

Figure 5.2.1.2 The area ratio of the two peaks from the real spectrum versus the ratio of the tentative model order L to the number of data points N. N=50. 104

Figure 5.2.1.3 Time domain amplitude of the 2-nd peak versus the ratio of the tentative model order to the number of data points. 105

Figure 5.2.1.4 Phase of the 2-nd peak versus the ratio of the tentative model order to the number of data points 106

Figure 5.2.1.5 Normalized damping rate of the 2-nd peak versus the ratio of the tentative model order to the number of data points. 107

Figure 5.2.1.6 Normalized frequency of the 2-nd peak versus the ratio of the tentative model order to the number of data points. 108

Figure 5.2.2.1 The area ratio of the two peaks from the real spectrum versus the ratio of the tentative model order L to the number of data points N. N=100. 111

Figure 5.2.2.2 Time domain amplitude of the 2-nd peak versus the ratio of the tentative model order to the number of data points. 112

Figure 5.2.2.3 Phase of the 2-nd peak versus the ratio of the tentative model order to the number of data points. 113

Figure 5.2.2.4 Normalized damping rate of the 2-nd peak versus the ratio of the tentative model order to the number of data points. 114

Figure 5.2.2.5 Normalized frequency of the 2-nd peak versus the ratio of the tentative model order to the number of data points. 115

Figure 5.2.3.1 Statistical result of the magnitude of the first three singular values of the LPSVD method for different numbers of data points N. ... 117

Figure 5.2.3.2 The area ratio of the two peaks from the real spectrum versus number of data points N. 118

Figure 5.2.3.3 Time domain amplitude of the 2-nd peak versus number of data points N. 119

Figure 5.2.3.4 Phase of the 2-nd peak versus number of data points N. 120

Figure 5.2.3.5 Normalized damping rate of the 2-nd peak versus number of data points N. 121

Figure 5.2.3.6 Normalized frequency of the 2-nd peak versus number of data points N. 122

Figure 5.3.1 A representative spectrum obtained from 512 points for section 5.3. 124

Figure 5.3.2 Statistical result of the magnitude of the first eight singular values of the LPSVD method. 124

Figure 5.3.3 Percentage frequency of finding different number of peaks versus the cut-off number of singular values of the LPSVD method for the example in section 5.3. 126

Figure 5.3.4 The area ratio of the 2-nd peak to the 1-st peak from the real spectrum versus the cut-off number of singular values of the LPSVD method........ 127

Figure 5.3.5 The area ratio of the 3-rd peak to the 1-st peak from the real spectrum versus the cut-off number of singular values of the LPSVD method........ 128

Figure 5.3.6 Time domain amplitude of the 1-st peak versus the cut-off number of singular values of the LPSVD method........ 129

Figure 5.3.7 Time domain amplitude of the 2-nd peak versus the cut-off number of singular values of the LPSVD method........ 130

Figure 5.3.8 Time domain amplitude of the 3-rd peak versus the cut-off number of singular values of the LPSVD method........ 131

Figure 5.3.9 Phase of the 1-st peak versus the cut-off number of singular values of the LPSVD method. 132

Figure 5.3.10 Phase of the 2-nd peak versus the cut-off number of singular values of the LPSVD method. 133

Figure 5.3.11 Phase of the 3-rd peak versus the cut-off number of singular values of the LPSVD method. 134

Figure 5.3.12 Normalized damping rate of the 1-st peak versus the cut-off number of singular values of the LPSVD method........ 135

Figure 5.3.13 Normalized damping rate of the 2-nd peak versus the cut-off number of singular values of the LPSVD method........ 136

Figure 5.3.14 Normalized damping rate of the 3-rd peak versus the cut-off number of singular values of the LPSVD method.......................... 137

Figure 5.3.15 Normalized frequency of the 1-st peak versus the cut-off number of singular values of the LPSVD method.................................. 138

Figure 5.3.16 Normalized frequency of the 2-nd peak versus the cut-off number of singular values of the LPSVD method.......................... 139

Figure 5.3.17 Normalized frequency of the 3-rd peak versus the cut-off number of singular values of the LPSVD method.......................... 140

Figure 5.4.1.1 Some representative DFT spectra for Example 5.4.1. 144

Figure 5.4.1.2 Statistical result of the first five singular values of the LPSVD method for the noise corrupted signals described in Example 5.4.1. 145

Figure 5.4.1.3 Magnitude of the first four singular values of the LPSVD method versus input time domain noise standard deviation σ_W. 145

Figure 5.4.1.4 The area ratio of the two peaks from the real spectrum versus A_2/σ_W, where A_2 is the time domain amplitude of the 2-nd peak and σ_W is time domain noise standard deviation.............................. 146

Figure 5.4.1.5 Time domain amplitude of the 1-st peak versus A_2/σ_W, where $A_2 = 250$. 147

Figure 5.4.1.6 Time domain amplitude of the 2-nd peak versus A_2/σ_W, where $A_2 = 250$.. 148

Figure 5.4.1.7 Phase of the 1-st peak versus A_2/σ_W, where $A_2 = 250$. 149

Figure 5.4.1.8 Phase of the 2-nd peak versus A_2/σ_W, where $A_2 = 250$. 150

Figure 5.4.1.9 Normalized damping rate of the 1-st peak versus A_2/σ_W, where $A_2 = 250$.. 151

Figure 5.4.1.10 Normalized damping rate of the 2-nd peak versus A_2/σ_W, where $A_2 = 250$ 152

Figure 5.4.1.11 Normalized frequency of the 1-st peak versus A_2/σ_W, where $A_2 = 250$ 153

Figure 5.4.1.12 Normalized frequency of the 2-nd peak versus A_2/σ_W, where $A_2 = 250$ 154

Figure 5.5.1.1 Some representative 1024 data points DFT spectra at different normalized frequency separations. 157

Figure 5.5.1.2 Statistical result of the first five singular values of the LPSVD method. Δf is the normalized frequency separation between the two peaks. 158

Figure 5.5.1.3 Magnitude of the first three singular values of the LPSVD method versus normalized frequency separation between two peaks. ... 158

Figure 5.5.1.4 The area ratio of the two peaks from the real spectrum versus the normalized frequency separation between the two peaks.......... 159

Figure 5.5.1.5 Time domain amplitude of the 1-st peak versus normalized frequency separation between the two peaks. 160

Figure 5.5.1.6 Time domain amplitude of the 2-nd peak versus normalized frequency separation between the two peaks. 161

Figure 5.5.1.7 Phase of the 1-st peak versus normalized frequency separation between the two peaks. 162

Figure 5.5.1.8 Phase of the 2-nd peak versus normalized frequency separation between the two peaks. 163

Figure 5.5.1.9 Normalized damping rate of the 1-st peak versus normalized frequency separation between the two peaks. 164

Figure 5.5.1.10 Normalized damping rate of the 2-nd peak versus normalized frequency separation between the two peaks. 165

Figure 5.5.1.11 Normalized frequency of the 1-st peak versus normalized frequency separation between the two peaks. 166

Figure 5.5.1.12 Normalized frequency of the 2-nd peak versus normalized frequency separation between the two peaks. 167

Figure 5.6.1.1 Some representative DFT spectra acquired using different normalized damping rates for the 2-nd peak. 170

Figure 5.6.1.2 Statistical result of the first five singular values of the LPSVD method. œ$_2$ is the normalized damping rate of the 2-nd peak. 171

Figure 5.6.1.3 Magnitude of the first three singular values of the LPSVD method versus normalized damping rate of the 2-nd peak. 171

Figure 5.6.1.4 The area ratio of the two peaks from the real spectrum versus input normalized damping rate of the 2-nd peak. 172

Figure 5.6.1.5 Time domain amplitude of the 1-st peak versus input normalized damping rate of the 2-nd peak. 173

Figure 5.6.1.6 Time domain amplitude of the 2-nd peak versus input normalized damping rate of the 2-nd peak. 174

Figure 5.6.1.7 Phase of the 1-st peak versus input normalized damping rate of the 2-nd peak. 175

Figure 5.6.1.8 Phase of the 2-nd peak versus input normalized damping rate of the 2-nd peak. 176

Figure 5.6.1.9 Normalized damping rate of the 1-st peak versus input normalized damping rate of the 2-nd peak. 177

Figure 5.6.1.10 Normalized damping rate of the 2-nd peak versus input normalized damping rate of the 2-nd peak. 178

Figure 5.6.1.11 Normalized frequency of the 1-st peak versus input normalized damping rate of the 2-nd peak. 179

Figure 5.6.1.12 Normalized frequency of the 2-nd peak versus input normalized damping rate of the 2-nd peak. 180

Figure 5.7.1.1 Some representative DFT spectra for different time domain amplitudes of the 2-nd peak. 183

Figure 5.7.1.2 Statistical result of the first eight singular values of the LPSVD method. A_2 is the input time domain amplitude of the 2-nd peak. 184

Figure 5.7.1.3 Magnitude of the first three singular values of the LPSVD method versus input time domain amplitude of the 2-nd peak. 184

Figure 5.7.1.4 The area ratio of the two peaks from the real spectrum versus input time domain amplitude of the 2-nd peak, A_2. 186

Figure 5.7.1.5 Time domain amplitude of the 1-st peak versus input A_2........... 187

Figure 5.7.1.6 Time domain amplitude of the 2-nd peak versus input A_2.......... 188

Figure 5.7.1.7 Phase of the 1-st peak versus input A_2.............................. 189

Figure 5.7.1.8 Phase of the 2-nd peak versus input A_2.............................. 190

Figure 5.7.1.9 Normalized damping rate of the 1-st peak versus input A_2......... 191

Figure 5.7.1.10 Normalized damping rate of the 2-nd peak versus input A_2. 192

Figure 5.7.1.11 Normalized frequency of the 1-st peak versus input A_2........... 193

Figure 5.7.1.12 Normalized frequency of the 2-nd peak versus input A_2. 194

Figure 5.8.1.1 Some representative DFT spectra for different time domain amplitudes of the 2-nd peak, transformed from 512 points FID data, which are composed of three decaying sinusoids plus Gaussian noise. 197

Figure 5.8.1.2 Statistical result of the first six singular values of the LPSVD method. A_2 is the input time domain amplitude of the 2-nd peak.......... 198

Figure 5.8.1.3 Magnitude of the first four singular values of the LPSVD method versus input time domain amplitude of the 2-nd peak, A_2........ 198

Figure 5.8.1.4 The area ratio of the 2-nd peak to the 1-st peak from the real spectrum versus input A_2. 199

Figure 5.8.1.5 The area ratio of the 3-rd peak to the 1-st peak from the real spectrum versus input A_2. 200

Figure 5.8.1.6 Time domain amplitude of the 1-st peak versus input A_2. 201

Figure 5.8.1.7 Time domain amplitude of the 2-nd peak versus input A_2.......... 202

Figure 5.8.1.8 Time domain amplitude of the 3-rd peak versus input A_2. 203

Figure 5.8.1.9 Phase of the 1-st peak versus input A_2. 204

Figure 5.8.1.10 Phase of the 2-nd peak versus input A_2. 205

Figure 5.8.1.11 Phase of the 3-rd peak versus input A_2. 206

Figure 5.8.1.12 Normalized damping rate of the 1-st peak versus input A_2. 207

Figure 5.8.1.13 Normalized damping rate of the 2-nd peak versus input A_2. 208

Figure 5.8.1.14 Normalized damping rate of the 3-rd peak versus input A_2. 209

Figure 5.8.1.15 Normalized frequency of the 1-st peak versus input A_2.......... 210

Figure 5.8.1.16 Normalized frequency of the 2-nd peak versus input A_2.......... 211

Figure 5.8.1.17 Normalized frequency of the 3-rd peak versus input A_2. 212

Figure 5.9.1.1 Some representative DFT spectra for different time domain amplitudes of the 3-rd peak, transformed from 512 points FID data which is composed of three decaying sinusoids plus Gaussian noise. ... 215

Figure 5.9.1.2 Statistical result of the first eight singular values of the LPSVD method. A_3 is the input time domain amplitude of the 3-rd peak. 216

Figure 5.9.1.3 Magnitude of the first three singular values of the LPSVD method versus input time domain amplitude of the 3-rd peak, A_3. 216

Figure 5.9.1.4 The area ratio of the 2-nd peak to the 1-st peak from the real spectrum versus input A_3. 217

Figure 5.9.1.5 Time domain amplitude of the 1-st peak versus input A_3.......... 218

Figure 5.9.1.6 Time domain amplitude of the 2-nd peak versus input A_3. 219

Figure 5.9.1.7 Phase of the 1-st peak versus input A_3................................ 220

Figure 5.9.1.8 Phase of the 2-nd peak versus input A_3.............................. 221

Figure 5.9.1.9 Normalized damping rate of the 1-st peak versus input A_3......... 222

Figure 5.9.1.10 Normalized damping rate of the 2-nd peak versus input A_3. 223

Figure 5.9.1.11 Normalized frequency of the 1-st peak versus input A_3.......... 224

Figure 5.9.1.12 Normalized frequency of the 2-nd peak versus input A_3......... 225

LIST OF SYMBOLS AND ABBREVIATIONS

a_k	complex time domain amplitude of the k-th component ($= A_k e^{j\phi_k}$)
f_k	frequency of the k-th component
h	Planck's constant (= 6.625×10^{-34} J-sec)
s(n)	noise-free complex damped exponential signals, unless otherwise stated
s_k	complex frequency of the k-th component ($=2\pi[-\alpha_k + jf_k]$)
ν_0	Larmor frequency ($\omega_0/(2\pi)$)
w(n)	zero mean white Gaussian noise signal with uncorrelated real and imaginary parts each having a variance of σ_w^2, unless otherwise stated
$\|\mathbf{x}\|_2$	2-norm of the vector $\mathbf{x}$ as defined in Equation (3.2.7)
y(n)	noise corrupted complex damped exponential signals
A_k	time domain amplitude of the k-th component
$\boldsymbol{A}^+$	pseudo-inverse of matrix $\boldsymbol{A}$
$\boldsymbol{A}^H$	complex conjugate transposition of matrix $\boldsymbol{A}$
$\boldsymbol{A}^T$	transposition of matrix $\boldsymbol{A}$
AR	autoregressive
$\mathbf{B}_0$	external static magnetic field
$\mathbf{B}_1$	external radio-frequency magnetic field
$\mathbf{B}_{eff}$	effective resultant field
CR	Cramer-Rao
DFT	discrete Fourier transformation
E( )	ensemble mean of the variable in the bracket
FID	free induction decay signal
$\boldsymbol{I}$	identity matrix
Im( )	imaginary part of the data in the bracket
LPSVD	Linear prediction singular value decomposition
M_{xy}	transverse magnetization
M_z	longitudinal magnetization
MLE	maximum likelihood estimation
MSE	mean-square error
NMR	nuclear magnetic resonance
PHD	Pisarenko harmonic decomposition
Re( )	real part of the data in the bracket

RF	radio-frequency
$\boldsymbol{R}_{ww}$	autocorrelation matrix of noise w(n)
$\boldsymbol{R}_{ss}$	autocorrelation matrix of signal s(n)
$\boldsymbol{R}_{yy}$	autocorrelation matrix of measured signal y(n) = s(n) + w(n), unless otherwise stated
SNR	time domain signal-to-noise ratio
$(\mathrm{SNR})_f$	frequency domain signal-to-noise ratio
SVD	singular value decomposition
TK	Tufts-Kumaresan
α_k	damping rate of the k-th component
ϕ_k	phase of the k-th component
γ	gyromagnetic (or magnetogyric) ratio
σ_w	time domain Gaussian noise standard deviation
$\sigma_k(\boldsymbol{A})$	k-th singular value of matrix $\boldsymbol{A}$
ω_k	$2\pi f_k$
ω_0	Larmor angular frequency ($=2\pi\nu_0$)
Δt	time domain sampling interval
œ_k	normalized damping rate of the k-th component ($= \alpha_k\Delta t$)
f_k	normalized frequency of the k-th component ($= f_k\Delta t$)
$\hat{}$	the corresponding variable is affected by noise or estimated under noise

Chapter 1

Introduction

1.1 Definition of the problem

The problem discussed here is that of estimating the unknown parameters of a time domain waveform y(n) that is the sum of M complex exponentially damped sinusoidal signals in white Gaussian noise. In particular, let there be given N samples of the waveform y(n), for n = 0, 1, ..., N-1, of a noise corrupted process,

$$
\begin{aligned}
y(n) &= \sum_{k=1}^{M} A_k e^{j\phi_k} e^{2\pi(-\alpha_k + jf_k)n\Delta t} + w(n) \quad , \\
&= \sum_{k=1}^{M} A_k e^{j\phi_k} e^{2\pi(-\text{œ}_k + j\mathit{f}_k)n} + w(n) \quad , \\
&= \sum_{k=1}^{M} a_k e^{s_k n} + w(n) \quad , \\
&= s(n) + w(n) \quad , \qquad 0 \le n \le N\text{-}1 \qquad (1.1.1)
\end{aligned}
$$

where A_k, ϕ_k, α_k and f_k are the amplitude, phase, damping rate and frequency of the k-th exponentially decaying signal, and the α_k's are positive for decaying signals, Δt is the sampling interval, and œ_k and f_k are defined as the normalized damping rate and normalized frequency, respectively, of the k-th exponentially decaying component. w(n) is a zero mean, white Gaussian noise signal with uncorrelated real and imaginary parts, each having a variance of σ_w^2. a_k and s_k, for k = 1, 2, . . . , M, are unknown complex numbers. M is also an unknown. In this thesis, we will denote ω_k, which is equal to $2\pi f_k$, as the angular frequency of the k-th exponentially decaying signal.

Given the N samples of noise corrupted data, the problem is to estimate the unknown value M, and also a_k, s_k, for $k = 1, 2, \ldots, M$.

1.2 Organization of the thesis

This thesis begins with an overview of the history of the above problem, some possible solutions and the objective of the thesis. In chapter 2, we will present a review of nuclear magnetic resonance (NMR) and the traditional ways of signal processing in NMR spectroscopy. Some mathematical background, namely, pseudo-inversion and singular value decomposition, that is essential to understand an advanced method called the linear prediction singular value decomposition (LPSVD) algorithm, is treated in chapter 3. In chapter 4, the LPSVD algorithm is treated in more detail, and a proposed algorithm described. In chapter 5, comparisons are made between the LPSVD algorithm, the proposed algorithm and a conventional NMR spectroscopy technique which uses the discrete Fourier transformation using simulated data. The estimation accuracy, measured in terms of bias, variance and mean square error of the parameter estimates of the three methods, is calculated based on several hundred independent trials and compared to the appropriate Cramer-Rao lower bounds. In chapter 6, major conclusions are listed.

1.3 History of the problem

The above is an age old problem and has occurred in many different areas. In pulsed nuclear magnetic resonance, to be described in more detail in Chapter 2, the signal collected, which is called the free induction decay (FID), is thought to be of the form of Equation (1.1.1). It is proportional to the net transverse magnetization in a frame rotating about the direction of a static magnetic field $\mathbf{B}_0$ with an angular velocity of $\omega_0 = \gamma B_0$, where γ is the gyromagnetic ratio of the nucleus being studied. This transverse magnetization is assumed to be decaying exponentially due to factors like transverse relaxation, magnetic field inhomogeneity, etc. Although the assumption of exponential decay is not always valid in NMR, for example, it is manifestly untrue in dipolar broadened solids, it is valid in liquid samples. Equation (1.1.1) is therefore expected to be an accurate description of the FID signal from tissue samples and to constitute a reasonable model for biomedical applications of NMR. By using a quadrature detection scheme for the NMR signal, and after analog-to-digital conversion, we will have a set of data composed of samples of a sum of M unknown, complex exponentially damped sinusoidal signals and additive noise described by Equation (1.1.1) (see section 2.6). In this thesis, we will compare some of the methods designed to estimate the parameters of such a model function.

1.4 Some current methods

Since this is an age old problem, an exhaustive review of the extensive literature is almost impossible. However, there are at least four major approaches to estimating the unknown parameters from the N given samples of data y(0), y(1), . . ., y(N-1). We shall briefly review these below.

In section 1.4.1 we discuss methods based on the maximum likelihood estimation (MLE). It is proven that for multiple sinusoids the MLE of the damping rates and frequencies is found by maximizing Equation (1.4.1.7), which unfortunately cannot be done analytically. However, if the sinusoidal frequencies are well resolved in a periodogram, and the damping rates approach zero, then the maximization of Equation (1.4.1.7) can be approximated by locating the peak frequencies and getting the full width at half-height of the peaks from the periodogram.

In section 1.4.2, the use of autoregressive (AR) spectral estimation for the sinusoidal damping rate and frequency estimation is shown to produce poor results at low signal-to-noise ratios (SNRs).

In an effort to improve the performance further, some methods based on principal component estimation and noise subspace estimation are described in sections 1.4.3 and 1.4.4. Principal component damping rate and frequency estimation methods discard the non-principal eigenvectors of the estimated autocorrelation matrix; Pisarenko harmonic decomposition, which is one of the methods based on the noise subspace eigenvector approach, finds the eigenvector of the autocorrelation matrix associated with the minimum eigenvalue and then roots a prediction error filter for the damping rates and frequencies as given by Equation (1.4.4.1.1).

1.4.1 Maximum likelihood estimation (MLE)

Assume that we are given N samples of data y(n) as described by Equation (1.1.1) where y(n) is composed of M complex damped sinusoids, and zero mean, complex, white Gaussian noise, w(n). Assume that the variance of w(n) is $2\sigma_w^2$.

The probability density function (PDF) of $\mathbf{y} = [\ y(0) \quad y(1) \ \ldots \ y(N-1)\ ]^T$

for a signal in complex Gaussian noise (Kay, 1987, p. 42) is

$$p(\mathbf{y} - \mathbf{s}) \;=\; \frac{1}{\pi^N \det(\boldsymbol{R}_{ww})} \exp\left[-(\,\mathbf{y} - \mathbf{s}\,)^H \boldsymbol{R}_{ww}^{-1} (\,\mathbf{y} - \mathbf{s}\,) \right] \qquad (1.4.1.1)$$

where 'H' denotes complex conjugate transposition, 'T' denotes transposition, $\boldsymbol{R}_{ww}$ is the NxN autocorrelation matrix of the noise and $\mathbf{s} = [\, s(0) \;\; s(1) \;\; \ldots\ldots \;\; s(N-1)\,]^T$ is the vector of signal samples, i.e., $s(n) = \sum_{k=1}^{M} a_k \, e^{\,s_k n}$.

Since the noise is assumed to be white,

$$\boldsymbol{R}_{ww} \;=\; 2\sigma_w{}^2 \boldsymbol{I},$$

where $\boldsymbol{I}$ is the identity matrix, and hence, to find the MLE of A_i, ϕ_i, œ_i and f_i, we need to maximize

$$p(\mathbf{y} - \mathbf{s}) \;=\; \frac{1}{\pi^N \det\left(2\sigma_w^2 \, \boldsymbol{I}\right)} \exp\left[-\frac{1}{2\sigma_w^2} (\,\mathbf{y} - \mathbf{s}\,)^H (\,\mathbf{y} - \mathbf{s}\,) \right] , \qquad (1.4.1.2)$$

for all A_i, ϕ_i, œ_i and f_i. Alternatively, maximizing Equation (1.4.1.2) is equivalent to minimizing

$$S \;=\; \left(\mathbf{y} - \sum_{i=1}^{M} a_i \, \mathbf{e}_{ci} \right)^H \left(\mathbf{y} - \sum_{i=1}^{M} a_i \, \mathbf{e}_{ci} \right) \qquad (1.4.1.3)$$

where $a_i \;=\; A_i \, e^{\,j\phi_i}$,

$\mathbf{e}_{ci} = [\; 1 \quad e^{s_i} \quad e^{2s_i} \quad \ldots \quad e^{(N-1)s_i} \;]^T$, and

$s_i \;=\; -2\pi \, \text{œ}_i + j\, 2\pi \, f_i$.

Let $\boldsymbol{E} \;=\; [\; \mathbf{e}_{c1} \;\; \mathbf{e}_{c2} \;\; \ldots\ldots \;\; \mathbf{e}_{cM} \;]$,

$\mathbf{a} \;=\; [\; a_1 \;\; a_2 \;\; \ldots\ldots \;\; a_M \;]^T$,

$\mathbf{f} \;=\; [\; f_1 \;\; f_2 \;\; \ldots\ldots \; f_M \;]^T$, and

$\mathbf{d} \;=\; [\; -\text{œ}_1 \;\; -\text{œ}_2 \;\; \ldots\ldots \;\; -\text{œ}_M \;]^T$, then Equation (1.4.1.3) can be written as

$$S(\mathbf{a}, \mathbf{d}, \mathbf{f}) = (\mathbf{y} - \boldsymbol{E}\mathbf{a})^H(\mathbf{y} - \boldsymbol{E}\mathbf{a}) . \qquad (1.4.1.4)$$

If $\boldsymbol{E}$ is assumed to be a known matrix, then $S(\mathbf{a}, \mathbf{d}, \mathbf{f})$ is minimized over $\mathbf{a}$ by

$$\hat{\mathbf{a}} = (\boldsymbol{E}^H\boldsymbol{E})^{-1}\boldsymbol{E}^H\mathbf{y} . \qquad (1.4.1.5)$$

Note that $\hat{\mathbf{a}}$ will be the MLE if $\boldsymbol{E}$ is replaced by $\hat{\boldsymbol{E}}$, the MLE of $\boldsymbol{E}$. Substituting Equation (1.4.1.5) into Equation (1.4.1.4) yields

$$\begin{aligned} S(\hat{\mathbf{a}}, \mathbf{d}, \mathbf{f}) &= \mathbf{y}^H(\mathbf{y} - \boldsymbol{E}\hat{\mathbf{a}}) \\ &= \mathbf{y}^H\mathbf{y} - \mathbf{y}^H\boldsymbol{E}(\boldsymbol{E}^H\boldsymbol{E})^{-1}\boldsymbol{E}^H\mathbf{y} . \qquad (1.4.1.6) \end{aligned}$$

To find the MLE of the damping rates and frequencies we must maximize

$$L(\mathbf{d}, \mathbf{f}) = \mathbf{y}^H\boldsymbol{E}(\boldsymbol{E}^H\boldsymbol{E})^{-1}\boldsymbol{E}^H\mathbf{y} , \qquad (1.4.1.7)$$

over $(\mathbf{d}, \mathbf{f})$. $L(\mathbf{d}, \mathbf{f})$ is a highly nonlinear function of the unknown normalized damping rates and frequencies. To maximize this function will require a search over 2M-dimensional space, which for several sinusoids is clearly impractical. Alternatively, an iterative optimization may be used, but convergence to the global maximum is not guaranteed. In fact, Bresler and Macovski (1986), and Kumaresan, et al. (1986) have developed iterative algorithms based on Equation (1.4.1.6).

If the MLE of the damping rates and frequencies can be obtained, then the MLE of the amplitudes and phases can easily be found from Equation (1.4.1.5).

To see how the periodogram is related to the MLE in a limited sense, consider the case of two sinusoids. Then,

$$\boldsymbol{E}^H\boldsymbol{E} = \begin{bmatrix} \mathbf{e}_{c1}^H\mathbf{e}_{c1} & \mathbf{e}_{c1}^H\mathbf{e}_{c2} \\ \mathbf{e}_{c2}^H\mathbf{e}_{c1} & \mathbf{e}_{c2}^H\mathbf{e}_{c2} \end{bmatrix}, \tag{1.4.1.8}$$

and so Equation (1.4.1.7) becomes

$$\mathrm{L}(\mathbf{d}, \mathbf{f}) = \frac{\mathbf{y}^H[\ \mathbf{e}_{c1} \quad \mathbf{e}_{c2}\ ]\begin{bmatrix} \mathbf{e}_{c2}^H\mathbf{e}_{c2} & -\mathbf{e}_{c1}^H\mathbf{e}_{c2} \\ -\mathbf{e}_{c2}^H\mathbf{e}_{c1} & \mathbf{e}_{c1}^H\mathbf{e}_{c1} \end{bmatrix}\begin{bmatrix} \mathbf{e}_{c1}^H \\ \mathbf{e}_{c2}^H \end{bmatrix}\mathbf{y}}{\left(\mathbf{e}_{c1}^H\mathbf{e}_{c1}\right)\left(\mathbf{e}_{c2}^H\mathbf{e}_{c2}\right) - \left|\mathbf{e}_{c1}^H\mathbf{e}_{c2}\right|^2} \tag{1.4.1.9}$$

$$= \frac{\left(\mathbf{e}_{c2}^H\mathbf{e}_{c2}\right)\left|\mathbf{e}_{c1}^H\mathbf{y}\right|^2 + \left(\mathbf{e}_{c1}^H\mathbf{e}_{c1}\right)\left|\mathbf{e}_{c2}^H\mathbf{y}\right|^2 - 2\mathrm{Re}\left(\mathbf{y}^H\mathbf{e}_{c1}\mathbf{e}_{c2}^H\mathbf{y}\ \beta\right)}{\left(\mathbf{e}_{c1}^H\mathbf{e}_{c1}\right)\left(\mathbf{e}_{c2}^H\mathbf{e}_{c2}\right) - |\beta|^2}, \tag{1.4.1.10}$$

where $\beta = \mathbf{e}_{c1}^H\mathbf{e}_{c2}$.

A sufficient condition for the MLE to equal the estimation from the periodogram given by Equation (1.4.1.11) is that $\beta = 0$, which is exactly equivalent to having both normalized damping rates œ_1 and œ_2 equal to zero and their normalized frequency separation be any integer multiple (besides zero) of $\frac{1}{N}$. This means that from Equation (1.4.1.10), that if $\beta = 0$, we need only maximize

$$\mathrm{L}(\mathbf{d}, \mathbf{f}) = \frac{1}{N}\left(|\mathrm{Y}(f_1)|^2 + |\mathrm{Y}(f_2)|^2\right), \tag{1.4.1.11}$$

where $\mathrm{Y}(f_i) = \sum_{n=0}^{N-1} y(n)\exp(-j2\pi f_i n)$, by choosing the two normalized frequencies that yield the two largest values of the periodogram.

For the case where both sinusoids are non-decaying, but their frequencies are not separated by an integer multiple of $\frac{1}{N}$, we can see from Figure 1.4.1 that $|\beta|$ is approximately equal to zero if $|f_1 - f_2| \gg \frac{1}{N}$. In other words, if the two peaks are well separated, the MLE of the frequencies may be approximated by finding the locations of the peaks from the periodogram.

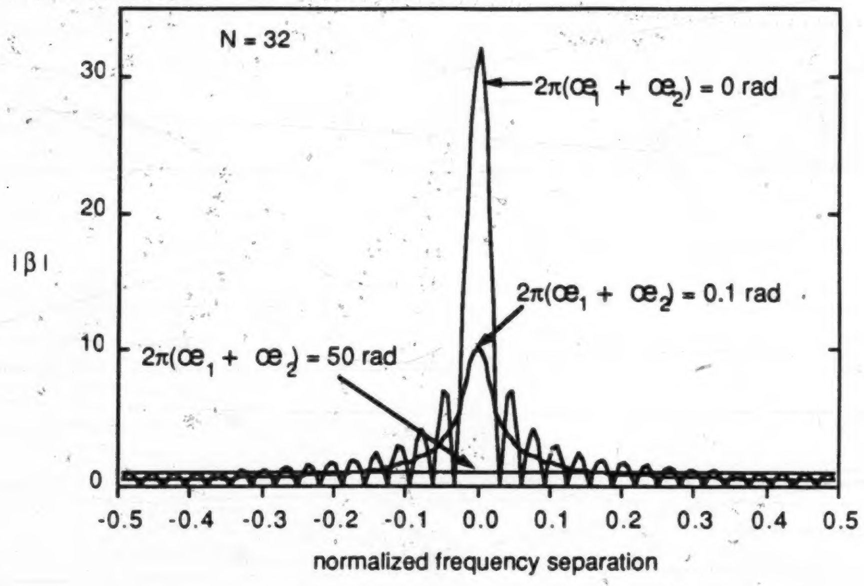


Figure 1.4.1 *$|\beta| = |e^H_{c1} e_{c2}|$ verses normalized frequency separation of two peaks for three different sets of $œ_1 + œ_2$, the sum of the two normalized damping rates.*

For the case where both sinusoids are decaying slightly, that is, $œ_1 \approx œ_2 \approx 0$, then

$$\mathbf{e}^H_{c1}\mathbf{e}_{c1} \approx \mathbf{e}^H_{c2}\mathbf{e}_{c2} \approx N,$$

and $|\beta|$ is approximately equal to zero provided that either $|f_1 - f_2| = \frac{m}{N}$, for m = 1, 2, . . . , $\frac{N}{2}$, or $|f_1 - f_2| >> \frac{1}{N}$ is satisfied, so that from Equation (1.4.1.10),

$$L(\mathbf{d}, \mathbf{f}) \approx \frac{1}{N}(|Y(f_1)|^2 + |Y(f_2)|^2). \tag{1.4.1.12}$$

The trend of $|\beta|$ for decreasing values of $œ_1 + œ_2$ can be seen from Figure 1.4.1.

In all of the above cases, the maximization of Equation (1.4.1.10) may be approximated by locating the peak frequencies and getting the full width at half-height of the peaks from the periodogram. For closely spaced or highly damped sinusoids, however, which is often the case in NMR, the periodogram is not close to the MLE. We must therefore maximize the exact expression as given by Equation (1.4.1.7) to obtain the MLE.

1.4.2 Damping rate and frequency estimation by spectral estimation

Since the periodogram cannot be used to estimate frequencies and damping rates of sinusoids that are not well resolved or are not close to non-decaying, much effort has been directed toward using the "high resolution" techniques of spectral estimation. But this approach may not be optimal in the maximum likelihood (ML) sense, because for damped sinusoids closely spaced in frequency, the MLE is not based on anything that may be interpreted as a spectral estimator. However, in light of the connection between the MLE and the spectral estimation of well resolved sinusoids as described in section 1.4.1, it is a reasonable alternative to the intractable MLE. In the following section, we choose to describe Prony's method because many indirect approaches to the problem are based on this idea, and this method is identical to the well known covariance method of linear prediction, which is one of the autoregressive (AR) spectral estimators.

1.4.2.1 Prony's method and some close relatives

Prony's method transforms the nonlinear estimation problem into a linear one. It is based on the following observation. Given N samples of s(n) which are composed of M complex damped exponential signals, as described by

$$s(n) = \sum_{k=1}^{M} a_k z_k^n , \qquad 0 \le n \le N-1 \qquad (1.4.2.2.1)$$

where $a_k = A_k e^{j\phi_k}$, $z_k = e^{s_k}$, and $s_k = -2\pi\alpha_k + j2\pi f_k$, it can then be shown (Oppenheim and Schafer, 1975) that s(n) may be generated by the recursive difference equation

$$s(n) = -\sum_{k=1}^{M} g(k)\, s(n-k), \qquad n \ge M, \qquad (1.4.2.2.2)$$

with the appropriate initial conditions $\{s(0), s(1), \ldots, s(M-1)\}$. Therefore, the signal may be viewed as the natural response of an autoregressive (or called all-pole) filter with the given initial conditions. The coefficients g(k) are related to the poles z_k by

$$G(z) = 1 + \sum_{k=1}^{M} g(k)\, z^{-k} = \prod_{k=1}^{M} (1 - z_k z^{-1}) = 0. \qquad (1.4.2.2.3)$$

Therefore, if we are given the data $\{s(0), s(1), \ldots, s(M-1)\}$, the poles may be determined exactly by solving the set of linear equations given by Equation (1.4.2.2.2) for n = M, M+1, . . ., 2M-1 to find the g(k)'s and then rooting the polynomial of Equation (1.4.2.2.3).

Summarizing, Prony's contribution was to transform the problem of finding the parameters that enter in a non-linear fashion, namely, s_k, k = 1, 2, . . . , M, into a problem of finding the coefficients g(1), g(2), . . . , g(M) of an M-th degree polynomial given by Equation (1.4.2.2.3).

To extend Prony's method to the case of exponential signals in noise, consider the following. If y(n) denotes the noise corrupted process and w(n) denotes white observation noise so that

$$y(n) \quad = \quad s(n) + w(n),$$

then Equation (1.4.2.2.2) becomes

$$y(n) - w(n) \quad = \quad - \sum_{k=1}^{M} g(k) \, [\, y(n-k) - w(n-k) \,], \qquad n \geq M,$$

or

$$y(n) \quad = \quad - \sum_{k=1}^{M} g(k) \, y(n-k) \; + \; w(n) + \sum_{k=1}^{M} g(k) \, w(n-k) \, . \qquad (1.4.2.2.4)$$

If we now define an error function $\varepsilon(n)$ such that,

$$\varepsilon(n) \quad = \quad w(n) + \sum_{k=1}^{M} g(k) \, w(n-k) \; , \qquad (1.4.2.2.5)$$

which may be thought of as an error due to noise, a least squares estimator of the g(k)'s is found by minimizing the sum of squares, E, i.e.,

$$\begin{aligned} E \quad &= \quad \frac{1}{N-M} \sum_{n=M}^{N-1} | \, \varepsilon(n) \, |^2 \\ &= \quad \frac{1}{N-M} \sum_{n=M}^{N-1} | \, y(n) + \sum_{k=1}^{M} g(k) \, y(n-k) \, |^2 \, , \qquad (1.4.2.2.6) \end{aligned}$$

with respect to the g(k)'s.

Notice that the minimization of Equation (1.4.2.2.6) is identical to that of the well known covariance method of linear prediction (Makhoul, 1975).

The minimization of Equation (1.4.2.2.6) is equivalent to solving

$$\begin{bmatrix} c_{yy}(1,1) & c_{yy}(1,2) & \cdots & c_{yy}(1,M) \\ c_{yy}(2,1) & c_{yy}(2,2) & \cdots & c_{yy}(2,M) \\ \vdots & \vdots & & \vdots \\ c_{yy}(M,1) & c_{yy}(M,2) & \cdots & c_{yy}(M,M) \end{bmatrix} \begin{bmatrix} g(1) \\ g(2) \\ \vdots \\ g(M) \end{bmatrix} = - \begin{bmatrix} c_{yy}(1,0) \\ c_{yy}(2,0) \\ \vdots \\ c_{yy}(M,0) \end{bmatrix}$$

$$\hat{\boldsymbol{R}}_{yy}\, \mathbf{g} \quad = \quad - \hat{\mathbf{r}}_{yy} \qquad (1.4.2.2.7)$$

in a least squares sense for g(k), k = 1, 2, . . . , M, where

$$c_{yy}(j,k) \quad = \quad \frac{1}{N-M} \sum_{n=M}^{N-1} y^{*}(n-j)\, y(n-k)\,, \qquad (1.4.2.2.8)$$

and "*" denotes complex conjugation. $\hat{\boldsymbol{R}}_{yy}$ is an estimated autocorrelation matrix of y(n).

Once we obtain an estimate of the g(k)'s (denoted by $\hat{g}(k)$), the values of the $\hat{z}_k$'s can be found by rooting the polynomial of Equation (1.4.2.2.3), and then the values of $\hat{a}_k$'s can be found by solving the system of linear equations given by Equation (1.4.2.2.1).

Generally, Prony's method will produce an inferior 'fit' of the data to that of the direct least squares approach because Prony's method attempts to minimize an equation error, `\i\su(n = M,N-1, ) | ε(n) |\s\up6(2)` , instead of the true least squares error $\sum_{n=0}^{N-1} | y(n) - \hat{s}(n) |^2$ (Kumaresan, 1982). To alleviate this problem, Kumaresan (1982) suggested increasing the order of the AR model employed (i.e., M in Equation (1.4.2.2.4)) to L where $M < L \leq N-M$. This means that he was proposing to minimize

$$E = \frac{1}{N-L} \sum_{n=L}^{N-1} | \varepsilon(n) |^2$$

$$= \frac{1}{N-L} \sum_{n=L}^{N-1} | y(n) + \sum_{k=1}^{L} g(k)\, y(n-k) |^2, \qquad (1.4.2.2.9)$$

with respect to the g(k)'s.

The minimization of Equation (1.4.2.2.9) is equivalent to solving

$$\begin{bmatrix} c_{yy}(1,1) & c_{yy}(1,2) & \cdots & c_{yy}(1,L) \\ c_{yy}(2,1) & c_{yy}(2,2) & \cdots & c_{yy}(2,L) \\ \vdots & \vdots & & \vdots \\ c_{yy}(L,1) & c_{yy}(L,2) & \cdots & c_{yy}(L,L) \end{bmatrix} \begin{bmatrix} g(1) \\ g(2) \\ \vdots \\ g(L) \end{bmatrix} = - \begin{bmatrix} c_{yy}(1,0) \\ c_{yy}(2,0) \\ \vdots \\ c_{yy}(L,0) \end{bmatrix}$$

$$\hat{\boldsymbol{R}}_{yy}\,\mathbf{g} = -\,\hat{\mathbf{r}}_{yy} \qquad (1.4.2.2.10)$$

in a least squares sense for g(k), where k = 1, 2, . . . , L. But Kumaresan also observed that at low signal-to-noise ratios, SNRs, increasing the AR model order may result in spurious peaks. A second suggestion was therefore made, namely, to find an estimate of M called $\hat{M}$, and the best subset of size $\hat{M}$ that best explains the data. What he did was to increase the value of $\hat{M}$ and calculate the corresponding best least squares error until a significant drop in the rate of decrease of the error occurred. By using these modifications Kumaresan showed by an example that Prony's method could be used at relatively lower SNR than had previously been thought possible.

1.4.3 Principal component estimation

It has been shown that AR spectral estimation provides reliable frequency estimates for high enough SNR. At lower SNRs the model order needs to be increased to the point where spurious peaks become a problem. This limits the utility of the spectral estimates at lower SNR. However, by choosing M out of L exponentials (where L > M) that best explain the data results in better estimates and provides us with some interesting insights.

Assuming that the phases of the M exponentially damped sinusoids in Equation (1.1.1) are independent random variables, uniformly distributed over $[0,2\pi)$, one can show that the theoretical L x L autocorrelation matrix $\boldsymbol{R}_{yy}$, where L > M, is composed of the signal autocorrelation matrix $\boldsymbol{R}_{ss}$ and the noise autocorrelation matrix $\boldsymbol{R}_{ww}$ as follows :

$$\begin{aligned} \boldsymbol{R}_{yy} &= \boldsymbol{R}_{ss} + \boldsymbol{R}_{ww} \\ &= \sum_{i=1}^{M} A_i^2 \mathbf{e}_{ci} \mathbf{e}_{ci}^H + \boldsymbol{R}_{ww}, \end{aligned} \tag{1.4.3.1}$$

where
$$\begin{aligned} \mathbf{e}_{ci} &= [\, 1 \;\; e^{s_i} \;\; e^{2s_i} \;\; \ldots \;\; e^{(L-1)s_i} \,]^T, \\ s_i &= -2\pi\, \text{œ}_i + j\, 2\pi f_i, \text{ and} \\ \boldsymbol{R}_{ww} &= 2\sigma_w^2 \boldsymbol{I}. \end{aligned}$$

$\boldsymbol{R}_{ss}$ is not full rank since it is of dimension L x L and the rank is M only. The total autocorrelation matrix $\boldsymbol{R}_{yy}$ is, however, full rank due to the inclusion of the $2\sigma_w^2\boldsymbol{I}$ term. Since $\boldsymbol{R}_{ss}$ is Hermitian (see Appendix 1), it can be decomposed as follows :

$$\boldsymbol{R}_{ss} = \sum_{i=1}^{M} \lambda_i \mathbf{v}_i \mathbf{v}_i^H, \tag{1.4.3.2}$$

where the λ_i's are real eigenvalues corresponding to the eigenvectors $\mathbf{v}_i$'s, and the $\mathbf{v}_i$'s are orthonormal (Stewart, 1973, pp. 308). This is called *eigendecomposition* or *spectral decomposition* and is a special case of the more general decomposition called *singular value decomposition (SVD)* as will be described in Chapter 3. Therefore, the eigendecomposition of $\boldsymbol{R}_{yy}$ produces

$$\boldsymbol{R}_{yy} \;=\; \sum_{i=1}^{M} (\lambda_i + 2\sigma_w^2)\mathbf{v}_i\mathbf{v}_i^H \;+\; \sum_{i=M+1}^{L} 2\sigma_w^2\,\mathbf{v}_i\mathbf{v}_i^H, \qquad (1.4.3.3)$$

where $\{\mathbf{v}_{M+1}, \mathbf{v}_{M+2}, \ldots, \mathbf{v}_L\}$ span the noise subspace. Equation (1.4.3.3) relies on the decomposition of the identity matrix

$$\boldsymbol{I} \;=\; \sum_{i=1}^{L} \mathbf{v}_i\mathbf{v}_i^H .$$

Equation (1.4.3.3) clearly demonstrates that the effect of noise is to introduce noise eigenvectors with, hopefully, small eigenvalues into the signal autocorrelation matrix. These non-principal eigenvectors bias the autocorrelation matrix estimate by adding components that are not present in the noiseless case. Hence, to gain some noise immunity it is reasonable to retain only the principal eigenvector components in the estimate of the autocorrelation matrix.

Methods that use standard spectral estimators but impose the constraint on the autocorrelation matrix that it be of low rank to match the rank of the theoretical signal autocorrelation matrix are generally called *principal component techniques.* The Tufts-Kumaresan (TK) algorithm, also called the linear prediction singular value decomposition (LPSVD) algorithm, is in fact one of the principal component methods. We are going to describe this algorithm in more detail in Chapter 4.

1.4.4 Noise subspace estimation

The second general approach to frequency estimation, which also relies on an eigendecomposition of the autocorrelation matrix, is based on the orthogonality of the signal vectors to the noise eigenvectors. It can be seen from Equations (1.4.3.1) and (1.4.3.2) that the principal eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M\}$ span the same subspace as the signal vectors $\{\mathbf{e}_{c1}, \mathbf{e}_{c2}, \ldots, \mathbf{e}_{cM}\}$. As a consequence, we have

$$\mathbf{e}_{ci}^H \sum_{k=M+1}^{L} \lambda_k \mathbf{v}_k = 0, \qquad i = 1, 2, \ldots, M, \qquad (1.4.4.1)$$

where the λ_k's are some arbitrary constants.

Hence, the signal vectors are orthogonal to all vectors in the noise subspace. This property forms the basis for the algorithm called Pisarenko harmonic decomposition (PHD). The PHD is now discussed.

1.4.4.1 Pisarenko harmonic decomposition (PHD)

The original PHD algorithm (Pisarenko, 1973) was applied to non-decaying sinusoids and L was assigned to equal M + 1, so that the noise subspace has the dimension 1. Therefore, as proven in general in Appendix 2, if $\mathbf{v}_{M+1}$ is the eigenvector of the (M+1) x (M+1) autocorrelation matrix $\boldsymbol{R}_{yy}$ associated with its minimum eigenvalue, then the zeros of

$$\sum_{n=0}^{M} [\mathbf{v}_{M+1}]_{n+1} \, z^{-n} \qquad (1.4.4.1.1)$$

are at $z_i = \exp(j2\pi f_i)$, for $i = 1, 2, \ldots, M$. Unfortunately, this method generally has poor performance in the presence of noise.

Some extensions to the PHD method have been proposed. Instead of using L = M+1, one such method used L > M+1. As shown in Appendix 2, the minimum eigenvalue is still $2\sigma_w^2$, but it now has a multiplicity of L-M. One can therefore average the L-M smallest eigenvalues and use this, hopefully, to improve the estimate.

Another approach was given by Reddi (1979) and Kumaresan (1982). They found a vector that was a linear combination of the eigenvectors in noise subspace. They then proceeded with the solution by fixing the first element of the solution vector to 1 and minimized the 2-norm of the other elements in the vector. In other words, as given in Appendix 2, if **g** is a linear combination of the noise subspace eigenvectors,

$$\mathbf{g} = \sum_{i=M+1}^{L} \lambda_i \mathbf{v}_i$$

where $\mathbf{g} = [\ g(0) \quad g(1) \quad \ldots \quad g(L-1)\ ]^T$ and $\{\mathbf{v}_{M+1}, \mathbf{v}_{M+2}, \ldots, \mathbf{v}_L\}$ are noise subspace eigenvectors, Reddi and Kumaresan constrained $g(0) = 1$ and minimized

$$\sum_{i=1}^{L-1} | g(i) |^2$$

with respect to λ_i. This technique has been shown to out perform the original PHD method and appears to work well for undamped sinusoids (Kumaresan, 1982).

Recently, Shinnar and Eleff (1988) applied a modified PHD technique to the NMR decaying FID signal. They modified the algorithm to incorporate knowledge of the noise power into the mathematical model and calculated a weighted average of the eigenvectors in order to reduce the effects of perturbations secondary to random noise on the eigenvectors. They said that by using these modifications, the algorithm may be useful in low SNR experiments.

1.5 Objectives of the thesis

The objective of the present study is to compare three methods designed to obtain 'accurate' estimates of the unknown parameters described in Equation (1.1.1). In section 1.4.1, we have shown that the periodogram cannot be used to estimate frequencies and damping rates of sinusoids that are not well resolved or are not close to non-decaying, so much effort has been directed toward using the "high resolution" techniques of spectral estimation. But these approaches may not be optimal in the maximum likelihood (ML) sense, because for damped sinusoids closely spaced in frequency, the MLE is not based on anything that may be interpreted as a spectral estimator. However, in light of the connection between the MLE and the spectral estimation of well resolved sinusoids as described in section 1.4.1, it is a reasonable alternative to the intractable MLE. One of the high resolution techniques of spectral estimation is Prony's method. But, generally, Prony's method resulted in poor estimates for noisy data. Since the theoretical autocorrelation matrix of Equation (1.1.1) (assuming the phases of the M exponentially damped sinusoids are independent random variables uniformly distributed over $[0,2\pi)$) is composed of the signal autocorrelation matrix and the noise autocorrelation matrix, hence, to gain some noise immunity, it is reasonable to retain only the principal eigenvector components in the estimate of the autocorrelation matrix. The Tufts-Kumaresan (TK) algorithm, also called the linear prediction singular value decomposition (LPSVD) algorithm, is one of the methods that uses standard spectral estimators but imposes the constraint on the autocorrelation matrix that it be of low rank to match the rank of the theoretical signal autocorrelation matrix.

In this thesis, we implemented the LPSVD algorithm, studied more about the algorithm's properties, developed another algorithm based on the LPSVD method and, using simulated data, compared the performance of the above two methods with a conventional NMR spectroscopy technique which uses the discrete Fourier transformation.

Chapter 2

The basis of NMR

2.1 Introduction

To understand nuclear magnetic resonance (NMR) spectroscopy and the traditional way of analyzing NMR spectra, it is necessary to provide some background material. The contents of this chapter are intended to provide a simple summary of the above topics. Detailed treatments of NMR phenomenon and NMR spectroscopy may be found in Abragam (1961), Slichter (1963), Farrar and Becker (1971), Gadian (1984), Harris (1986), and Ernst et al. (1988).

Nuclei which possess spin angular momenta and therefore magnetic moments will, when allowed to equilibrate in a static magnetic field, distribute themselves among discrete stationary energy states. Transitions of the nuclear spin between one state and the next can be induced by the application of electromagnetic radiation at the appropriate frequency and this gives rise to the nuclear magnetic resonance phenomenon. As we shall see, the term resonance implies that the exciting radiation is in tune with a natural frequency of the magnetic system, in this case the frequency of gyroscopic precession of the magnetic moment in the external static magnetic field. A proper description of this phenomenon requires quantum mechanics. However, a number of properties are more easily visualized through a classical treatment (Slichter, 1963; Farrar & Becker, 1971).

2.2 Simple resonance theory

Nuclei are made up of nucleons called protons and neutrons. A nucleus with either an odd number of protons and/or neutrons has a nuclear spin angular momentum that is characterised by the total nuclear angular momentum quantum number, I, which is an integer or a half-integer greater than zero. In fact, the value of I for each nucleus depends on the number of protons and neutrons. This is summarized in the following table.

proton number	neutron number	spin number I
odd	odd	integer: 1, 2, 3, ...
odd	even	half integer: $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$
even	odd	half integer: $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$
even	even	0

Nuclei of interest in biological applications of NMR and having spin $\frac{1}{2}$ are ^{1}H, ^{13}C, ^{31}P, and ^{19}F.

A nucleus with unpaired nucleons (i.e., $I \geq \frac{1}{2}$) possesses a net angular momentum (total nuclear angular momentum)

$$\mathbf{J} = \mathbf{I}\frac{h}{2\pi} = \mathbf{I}\hbar \tag{2.2.1}$$

and, as a result, a net magnetic moment,

$$\boldsymbol{\mu} = \gamma \mathbf{J} = \gamma \hbar \mathbf{I} \tag{2.2.2}$$

colinear with it. Here the proportionality constant γ is called the gyromagnetic (or magnetogyric) ratio, a unique constant for a given nucleus; h is Planck's constant ($=6.625\times10^{-34}$ J-sec) and **I** is a dimensionless spin angular momentum operator.

The magnitude J of the total nuclear angular momentum is given in terms of I by the relation

$$J = \sqrt{I(I+1)}\,\hbar \ .$$

In the presence of a static magnetic field $\mathbf{B}_0$, the vector $\boldsymbol{\mu}$ can assume only certain discrete orientations with respect to $\mathbf{B}_0$. In this thesis, we assume that the direction of $\mathbf{B}_0$ is along the z-axis unless otherwise stated. The measurable components of $\boldsymbol{\mu}$ in the direction defined by $\mathbf{B}_0$ are given by the magnetic quantum number, m, in the equation

$$J_z = \hbar m \ ; \qquad m = I, I\text{-}1, I\text{-}2, \ldots, \text{-}I \ .$$

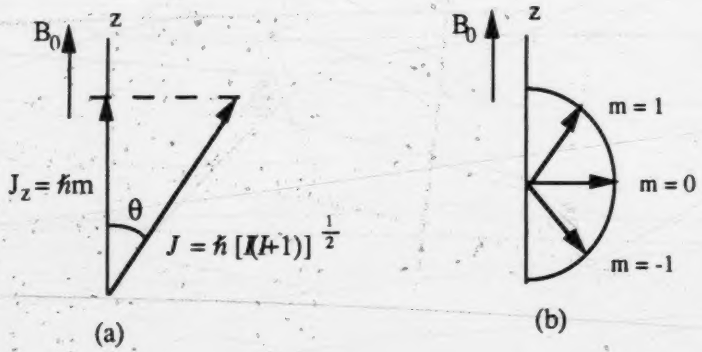


***Figure 2.2.1** Angular momentum vectors : (a) showing the use of quantum numbers I and m, (b) the possible orientations for an angular momentum vector of quantum number = 1.*

Hence, there are $2I+1$ such discrete orientations. These orientations correspond to different energy states as follows :

$$\begin{aligned} E &= -\boldsymbol{\mu} \cdot \mathbf{B}_0 \\ &= -\mu_z B_0 \\ &= -m \gamma \hbar B_0 , \qquad m = I, I\text{-}1, I\text{-}2, \ldots, \text{-}I \ , \end{aligned} \tag{2.2.3}$$

where μ_z ($= m\gamma\hbar$) is the z component of the nuclear magnetic moment. The energy differnece between two such states is referred to as "nuclear Zeeman splitting" of the energy levels. Energy levels are shown in Figure 2.2.2 for nuclei with spin quantum numbers $I = \frac{1}{2}$ and 1. The lowest energy state is defined as the one in which the measurable component of $\boldsymbol{\mu}$ aligns parallel to the field $\mathbf{B_0}$. The highest one corresponds to the case where $\boldsymbol{\mu}$ is anti-parallel to $\mathbf{B_0}$.

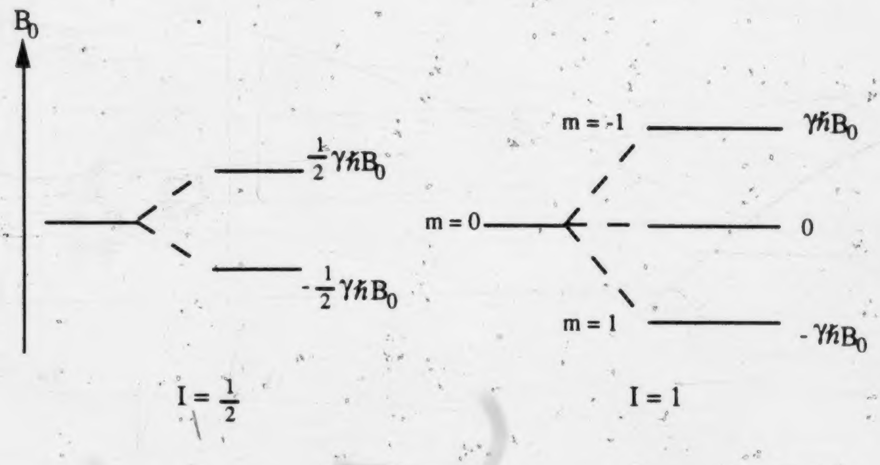


Figure 2.2.2 *Magnetic energy levels for nuclei of spin $\frac{1}{2}$ and 1.*

Restricting ourselves to spin-half ($I = \frac{1}{2}$) nuclei results in two energy states, an upper and a lower state denoted by E_{upper} and E_{lower}, respectively. The energy separation ΔE between these two states is

$$\Delta E \;=\; E_{upper} - E_{lower} \;=\; \gamma\hbar B_0 \;. \qquad (2.2.4)$$

The Bohr condition ($\Delta E = h\upsilon$) permits us to give the frequency of the radiation required to excite the nuclear transition, namely

$$\upsilon_0 = \frac{\Delta E}{h} = \frac{\gamma\hbar B_0}{h} = \frac{\gamma}{2\pi} B_0 \ , \qquad (2.2.5)$$

where υ_0 is called the Larmor frequency. Normally, magnetic fields in the range of 0.5 T to 14 T are used for biomedical NMR, and it can be seen that frequencies of the order of tens to hundreds of MHz are necessary. Therefore, for any particular nucleus in a given magnetic field, the NMR frequency will be characteristic, depending exclusively on the gyromagnetic ratio peculiar to that particular nucleus. Some examples are given in Harris (1986, p.5) and Bovey (1988, p.3). Transitions of the spins between these energy states, brought about through the application of an alternating electromagnetic field of frequency υ_0, is referred to as the "*on resonance*" condition.

In Figure 2.2.3, the separation of the proton magnetic energy levels is shown as a function of magnetic field strength.

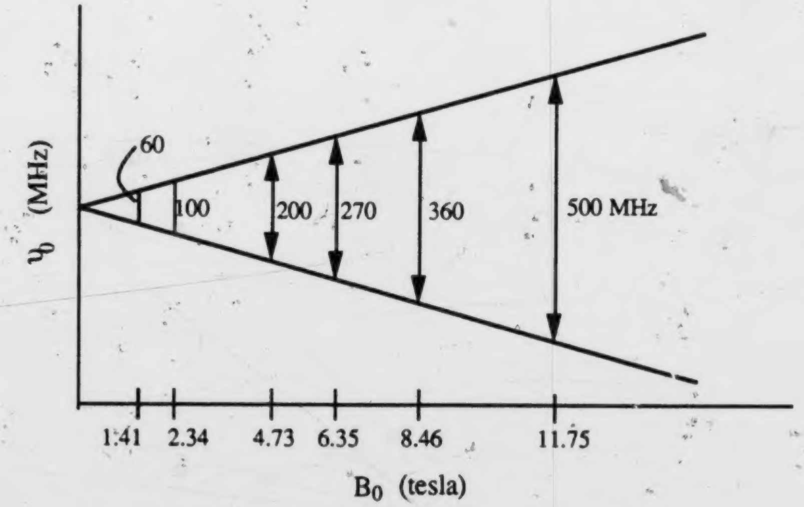


***Figure 2.2.3** The splitting of magnetic energy levels of protons, expressed as resonance frequency υ_0, as a function of magnetic field strength.*

2.3 Interaction between magnetization and magnetic field

In treating the experimental observation of nuclear magnetic resonance, it is convenient to use Bloch's approach. Bloch's approach was macroscopic and classical and is complementary to the use of a spin Hamiltonian. The Bloch equations can be used to describe the phenomena involved in a simple spectrum, but they are not applicable to a complex spectrum.

In the next subsection, 2.3.1, the effect of a static magnetic field on the magnetization will be discussed. It is shown that in a static magnetic field B_0, the magnetization will precess around the field with an angular velocity $\omega_0 = -\gamma B_0$. In subsection 2.3.2, we will show the influence of an applied radio-frequency field on the magnetization that was originally at thermal equilibrium. It will be shown that the magnetization will precess in a cone of fixed angle about an effective field, $\mathbf{B}_{eff}$, given by Equation (2.3.2.3) at an angular frequency of $-\gamma\mathbf{B}_{eff}$ in a frame rotating at the angular frequency of the applied radio-frequency field.

First, we define the total magnetic moment per unit volume of sample or magnetization **M**. The magnetization is the resultant sum, per unit volume, of all the individual magnetic moments with the magnetogyric ratio γ, $\boldsymbol{\mu}$, in an assembly of identical nuclei,

$$\mathbf{M} = \sum_j \boldsymbol{\mu}_j \ , \qquad (2.3.1)$$

or in terms of the total spin angular momentum **J**,

$$\mathbf{M} = \gamma \mathbf{J} \ . \qquad (2.3.2)$$

In the absence of a static field $\mathbf{B}_0$, the magnetic moments are oriented at random. However, when $\mathbf{B}_0$ is applied more of the magnetic moments orient along the field than against it as predicted in Equation (2.5.1.2). Therefore, a net magnetization of amplitude M_0 may be assigned along the axis colinear with $\mathbf{B}_0$. Furthermore, since the orientation of the xy component of each magnetic moment is uncorrelated at equilibrium, the average xy component of the resultant equilibrium magnetization M_{xy} is zero.

2.3.1 Motion in a static magnetic field : Larmor precession

If **M** is perturbed away from the z-axis it precesses around $\mathbf{B}_0$ with an angular frequency $\omega_0 = -\gamma\mathbf{B}_0$ as seen in the laboratory frame of reference. This can be shown as follows. Consider an ensemble of magnetic moments placed in a static magnetic field $\mathbf{B}_0$. The effect of an applied magnetic field $\mathbf{B}_0$ can be predicted from classical mechanics. The field results in a torque on the magnetization, changing the angular momentum **J** according to

$$\frac{d}{dt}\mathbf{J} = \mathbf{M} \times \mathbf{B}_0 \ ,$$

or

$$\frac{d}{dt}\mathbf{M}(t) = \gamma\mathbf{M}(t) \times \mathbf{B}_0 \ . \qquad (2.3.1.1)$$

In the absence of both relaxation effects and rotating fields $\mathbf{B}_1$, the projection of $\mathbf{M}(t)$ on the z-axis, M_z, would remain constant, i.e.,

$$\frac{d}{dt}M_z(t) = 0 \ . \qquad (2.3.1.2)$$

The magnitudes of the x and y projections, $M_x(t)$ and $M_y(t)$, will vary with time as $\mathbf{M}(t)$ precesses, and can be expressed by

$$\left.\begin{aligned} \frac{d}{dt}M_x(t) &= \gamma(\mathbf{M}(t)\times\mathbf{B}_0)_x = \gamma M_y B_0 = \omega_0 M_y \\ \frac{d}{dt}M_y(t) &= \gamma(\mathbf{M}(t)\times\mathbf{B}_0)_y = -\gamma M_x B_0 = -\omega_0 M_x \end{aligned}\right\} \qquad (2.3.1.3)$$

In other words, $\mathbf{M}(t)$ rotates about $\mathbf{B}_0$ with an angular velocity of $\boldsymbol{\omega}_0 = -\gamma\mathbf{B}_0$, the Larmor angular frequency, in addition to having a constant magnitude and a constant angle with $\mathbf{B}_0$ (see Figure 2.3.1.1).

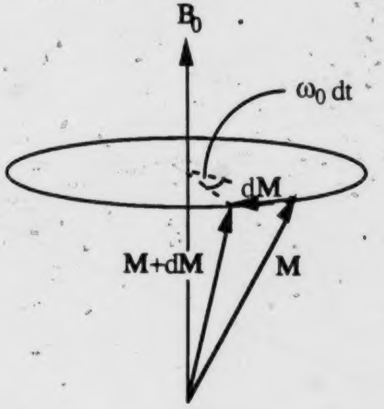


Figure 2.3.1.1 *Precession of a classical angular momentum* $\boldsymbol{J}$, *having a proportional magnetization* $\boldsymbol{M} = \gamma\boldsymbol{J}$, *about a static magnetic field* $\boldsymbol{B}_0$.

2.3.2 The influence of a rotating field : resonance

One way to perturb the net magnetization from its thermal equilibrium state is by irradiating the spin system with a circularly polarized RF magnetic field of angular frequency $\boldsymbol{\omega}_{rf}$ and amplitude B_1 restricted to the transverse plane and rotating in the same sense as the spins. Let us now look at the effect of the RF field.

The equation of motion of **M**(t) in the presence of the total field

$$\mathbf{B}(t) = \mathbf{B}_0 + \mathbf{B}_1(t)$$

becomes

$$\frac{d}{dt}\mathbf{M}(t) \;=\; \mathbf{M}(t) \times \gamma[\,\mathbf{B}_0 + \mathbf{B}_1(t)\,]\;. \tag{2.3.2.1}$$

To solve this equation, it is convenient to place ourselves in a frame rotating about the direction of $\mathbf{B}_0$ with the same angular velocity as $\mathbf{B}_1(t)$. In Figure 2.3.2.1, we shall designate by Oxyz (unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$) a fixed laboratory coordinate system whose Oz axis is the direction of the field $\mathbf{B}_0$, and by Ox'y'z' (unit vectors $\mathbf{e}_{x'}$, $\mathbf{e}_{y'}$, $\mathbf{e}_{z'}$) the axes obtained from Oxyz by rotation through an angle $\omega_{rf}t$ about Oz, where Ox' is the direction of the rotating field $\mathbf{B}_1(t)$. Then, the relative velocity of the vector **M**(t) in the rotating frame is

$$\begin{aligned}
\frac{d}{dt}\mathbf{M}^r(t) &= \frac{d}{dt}\mathbf{M}(t) \;-\; \boldsymbol{\omega}_{rf} \times \mathbf{M}(t) \\
&= \mathbf{M}(t) \times \gamma\,\mathbf{B}(t) \;-\; \boldsymbol{\omega}_{rf} \times \mathbf{M}(t) \\
&= \mathbf{M}(t) \times \gamma\,\mathbf{B}_{eff}(t)\;,
\end{aligned} \tag{2.3.2.2}$$

where $\mathbf{M}^r(t)$ is the magnetization vector in the rotating frame, and the "*effective field*" $\mathbf{B}_{eff}(t)$ (which is static with respect to the rotating reference frame) can be written as

$$\begin{aligned}
\mathbf{B}_{eff}(t) &= \mathbf{B}(t) + \frac{\boldsymbol{\omega}_{rf}}{\gamma} \\
&= \mathbf{B}_0 + \mathbf{B}_1(t) + \frac{\boldsymbol{\omega}_{rf}}{\gamma} \\
&= B_1(t)\,\mathbf{e}_{x'} + \left(B_0 - \frac{\omega_{rf}}{\gamma} \right) \mathbf{e}_{z'}\;.
\end{aligned} \tag{2.3.2.3}$$

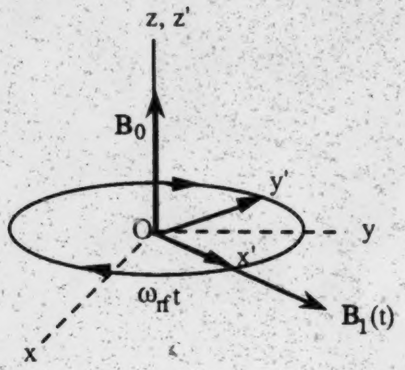


Figure 2.3.2.1 Oxyz is a fixed laboratory coordinate system. The static magnetic field B_0 is along the Oz axis. The Ox'y'z' is a rotating system about Oz with the angular frequency ω_{rf}, and Ox' axis is the direction of the rotating field $B_1(t)$.

Physically, Equations (2.3.2.2) and (2.3.2.3) state that in the rotating frame the magnetization acts as though it effectively experiences a static magnetic field $\mathbf{B}_{eff}(t)$. In the rotating frame therefore, the magnetization precesses in a cone of fixed angle about the direction of $\mathbf{B}_{eff}(t)$ at angular frequency $-\gamma B_{eff}(t)$. The situation is illustrated in Figure 2.3.2.2 for a magnetization which, at t = 0, was orientated along the z-direction. The angle β by which $\mathbf{M}$ tilts away from the z'-axis about $\mathbf{B}_{eff}$ in the clockwise direction under the action of the RF magnetic field $\mathbf{B}_{eff}$ of duration τ_p (commonly known as the pulse width) is

$$\beta \quad = \quad \gamma B_1 \tau_p . \qquad (2.3.2.4)$$

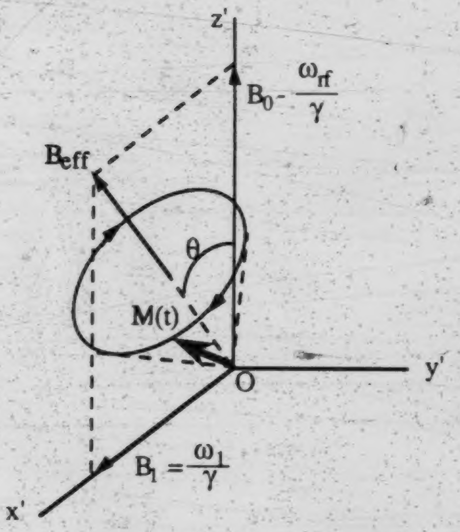


Figure 2.3.2.2 In the rotating frame Ox'y'z', the effective field $\boldsymbol{B}_{eff}$ *has a fixed direction about which the magnetic moment* ***M(t)*** *rotates with a constant angular frequency* $-\gamma \boldsymbol{B}_{eff}$ *about* $\boldsymbol{B}_{eff}$

Consider the special case where the angular frequency of $\mathbf{B}_1$ is the same as $\boldsymbol{\omega}_0 = -\gamma\mathbf{B}_0$. This is called the "on resonance" condition. From Equation (2.3.2.3), the effective field is then simply $B_1\mathbf{e}_{x'}$. Therefore, magnetization that is parallel to the static field initially will precess in the y'z' plane at a frequency $\boldsymbol{\omega}_1 = -\gamma\mathbf{B}_1$ for as long as the RF irradiation persists. A short application of an RF magnetic field, $\mathbf{B}_1$, which tilts **M** by exactly 90° or 180° is referred to as a 90° or a 180° pulse, respectively. Following a 90° pulse, the excess magnetization will be perpendicular to $\mathbf{B}_0$ and will precess at angular frequency $-\gamma\mathbf{B}_0$ in the laboratory frame, as shown in Figure 2.3.2.3.

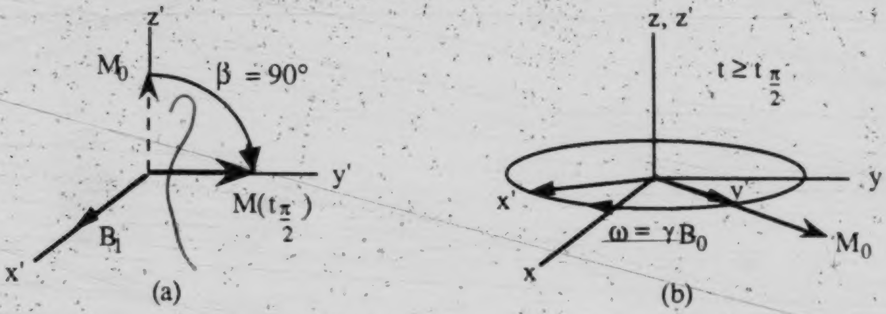


*Figure 2.3.2.3 (a) Nutation of **M** away from its equilibrium position onto the rotating frame y' axis produced by a 90° RF pulse at resonance polarized along +x' axis of rotating frame. (b) Continued free precession of **M** in the laboratory frame after the termination of the RF pulse.*

2.4 The chemical shift

The resonance frequency of a nucleus is directly proportional to the local magnetic field experienced by the nucleus. Since the applied field $\mathbf{B_0}$ induces electronic currents in atoms and molecules, and these produce an additional small field $\mathbf{B_{0\sigma}}$ at the nucleus proportional to $\mathbf{B_0}$, not all nuclei in a sample may experience the same local magnetic field. The total effective local field $\mathbf{B_{loc}}$ at the nucleus can therefore be written

$$\begin{aligned} \mathbf{B_{loc}} &= \mathbf{B_0} - \mathbf{B_{0\sigma}} \\ &= \mathbf{B_0}\,(1 - \sigma)\,, \end{aligned} \tag{2.4.1}$$

where σ is a dimensional constant, known as the shielding or screening constant. It has values typically in the region 10^{-6} to 10^{-3} (Gadian, 1984, p.5). σ is dependent upon the electronic environment of the nucleus, and therefore nuclei in different chemical environments experience different fields and hence produce signals at different frequencies. The separation of resonance frequencies from an arbitrarily chosen reference frequency is termed the chemical shift, and is expressed in terms of the dimensionless units of parts per million (ppm). If one nucleus is more shielded than another, its signal will be shifted to a lower frequency; by convention it will appear further towards the right-hand side of the spectrum. Table 2.4.1 indicates some of the chemical shifts observed in NMR spectroscopy, along with some values observed in *in-vivo* NMR spectroscopy (Jardetzky and Roberts, 1981, p.27).

nucleus	chemical shift range (ppm) in high-resolution spectra	approximate chemical shift range (ppm) in *in-vivo* spectra
^{1}H	12	8
^{13}C	350	200
^{31}P	700	30

Table 2.4.1 *Chemical shifts observed in high-resolution and in* in-vivo *spectroscopy.*

2.5 Relaxation phenomena

As indicated in the last section, a 90° RF pulse rotates the magnetization **M** into the x'y' plane (i.e., $M_z = 0$). However, after termination of the RF pulse, we know from experience that the net magnetization will return to the z-axis, and the transverse component of the magnetization will disappear. They are called longitudinal (or spin-lattice) relaxation and transverse (or spin-spin) relaxation, respectively. In the following subsections, we will briefly describe these phenomena.

2.5.1 Longitudinal relaxation

Just after a 90° RF pulse, the net magnetization **M** lies in the x'y' plane. This means that the proportion of the spins in the lower energy state is the same as that in the upper energy state. However, the eventual return of **M** to the z-axis after termination of the RF pulse means that in returning to equilibrium, the populations readjust so that there are more spins in the lower (spin-up) energy state than in the upper (spin-down) energy state. Hence, there must exist some means by which the upper level spins could relax to the lower level in order to reestablish an equilibrium spin population distribution and permit a resonance signal to be observed again.

This mechanism can occur because each spin is not entirely isolated from the rest of the assembly of atoms, commonly referred to as the "lattice". The spins and the lattice may be considered to be essentially separate coexisting systems with a very inefficient but nevertheless very important link by which thermal energy may be exchanged. This link is provided by the modulation of spin interactions by "molecular motion". Each nucleus sees a number of other nearby magnetic nuclei, both in the same molecule and in other molecules. These neighboring nuclei are in motion with respect to the observed nucleus, and because these neighboring nuclei have magnetic moments, this motion gives rise to fluctuating magnetic fields. If these internal fluctuating fields have frequency components at the resonant (Larmor) frequency, they can stimulate the de-excitation of nuclear spins from the upper state, the energy being transferred from the spin system into thermal kinetic energy of the lattice (or heat reservoir). Let us denote the probability per second that such a coupling will induce a spin transition upward in energy (from lower energy level to upper level) by $W_\uparrow$, and the reverse process by $W_\downarrow$. If the number of spins in the lower and upper energy state are denoted by $N_{lower}(t)$ and $N_{upper}(t)$, respectively, then we have a rate equation

$$\frac{d}{dt} N_{lower}(t) \quad = \quad N_{upper}(t)W_{\downarrow} - N_{lower}(t)W_{\uparrow} \ . \tag{2.5.1.1}$$

Since at equilibrium $\frac{d}{dt} N_{lower}(t)$ is zero, Equation (2.5.1.1) tells us that

$$\frac{N^0_{upper}}{N^0_{lower}} \quad = \quad \frac{W_{\uparrow}}{W_{\downarrow}}$$

$$= \quad e^{-\Delta E / kT} \quad = \quad e^{-\gamma\hbar B_0 / kT} \ , \tag{2.5.1.2}$$

where N^0_{upper} and N^0_{upper} are the number of spins in the upper and lower energy state, respectively, at equilibrium. The last two equalities in Equation (2.5.1.2) assume that at equilibrium the relative spin populations are given by the Boltzmann distribution, where k is Boltzmann's constant which is equal to 1.381 x 10^{-34} J/°K.

Let $n(t) = N_{lower}(t) - N_{upper}(t)$ be the difference in population of the energy levels, n_0 be the equilibrium population difference, and N be the total number of spins. The two variables $N_{lower}(t)$ and $N_{upper}(t)$ may be replaced by n(t) and N, using the equations

$$N \quad = \quad N_{lower}(t) + N_{upper}(t)$$

$$n\,(t) \quad = \quad N_{lower}(t) - N_{upper}(t) \ .$$

This results in

$$\left.\begin{aligned} N_{lower}(t) &= \frac{1}{2}(N+n(t)) \\ N_{upper}(t) &= \frac{1}{2}(N-n(t)) \end{aligned}\right\} \tag{2.5.1.3}$$

Substitution of Equation (2.5.1.3) into Equation (2.5.1.1) gives us

$$\frac{d}{dt} n(t) \quad = \quad N\,(\,W_{\downarrow} - W_{\uparrow}\,) - n(t)\,(\,W_{\downarrow} + W_{\uparrow}\,) \ . \tag{2.5.1.4}$$

Hence, at equilibrium we have :

$$\frac{n_0}{N} = \frac{W_\downarrow - W_\uparrow}{W_\downarrow + W_\uparrow}$$

and so, defining

$$\frac{1}{T_1} = W_\downarrow + W_\uparrow$$

we may rewrite Equation (2.5.1.4) as:

$$\frac{d}{dt} n(t) = \frac{1}{T_1} (n_0 - n(t)) . \qquad (2.5.1.5)$$

By analogy to Equation (2.5.1.5), and recognizing that $M_z = \mu_z n$, Bloch expected that it would be reasonable for M_z to be established according to the equation

$$\frac{d}{dt} M_z(t) = - \frac{1}{T_1} (M_z(t) - M_0) \qquad (2.5.1.6)$$

where M_0 is the thermal equilibrium magnetization. In other words,

$$M_z(t) = M_0 + (M_z(0_+) - M_0) e^{-t/T_1} , \qquad (2.5.1.7)$$

where $M_z(0_+)$ is the z component of the magnetization just after the RF pulse. Hence Equation (2.5.1.7) predicts an exponential return to equilibrium with time constant T_1. Following a 90° RF pulse, $M_z(0_+) = 0$, hence

$$M_z(t) = M_0 (1 - e^{-t/T_1}) . \qquad (2.5.1.8)$$

T_1 is therefore a measure of the time required to establish thermal equilibrium between the spins and their surroundings. If a static field $\mathbf{B_0}$ is applied to a sample that is initially unmagnetized, T_1 is also the characteristic time for the equilibrium magnetization $\mathbf{M}_0$ to be established. T_1 is called the *spin-lattice relaxation time* or the *longitudinal relaxation time*, emphasizing the relationship of M_z and the static field.

The nature of the rotational and translational motions permitted by the lattice significantly influences the longitudinal relaxation efficiency. Remember that it is only the component of the fluctuating magnetic fields at the Larmor frequency which is efficient in inducing T_1 relaxation. Although the time behavior of the thermal motion is random, it can be Fourier analyzed to estimate its component at the Larmor frequency. A method of analysis of this random motion is the following. The randomness of the thermal motion produces random magnetic fields over all time and space. The magnitude of these fluctuating fields can be characterized by the autocorrelation function which measures how well the fluctuating field correlates with itself at a certain time later. If the process is wide sense stationary (i.e., the autocorrelation function is a function only of the time difference) and if the autocorrelation function is absolutely integrable, then the power spectral density function exists for all frequencies, and is given by the Fourier transform of the autocorrelation function. The power spectral density function tells us the distribution of power over all frequencies. Assuming that the autocorrelation function is simply a mono-exponential decaying function (therefore absolutely integrable) with the correlation time τ_c (a measure of the average time that the environment can be considered to be in a particular state), then the power spectral density function becomes

$$J(\omega) = G(0) \frac{\tau_c}{1+\omega^2\tau_c^2} ,$$

where G(0) is the autocorrelation function at time $t = 0$. The forms of $J(\omega)$, where we assume $\int_0^\infty J(\omega)\, d\omega = \text{constant}$ for different τ_c values, are shown in Figure 2.5.1.1.

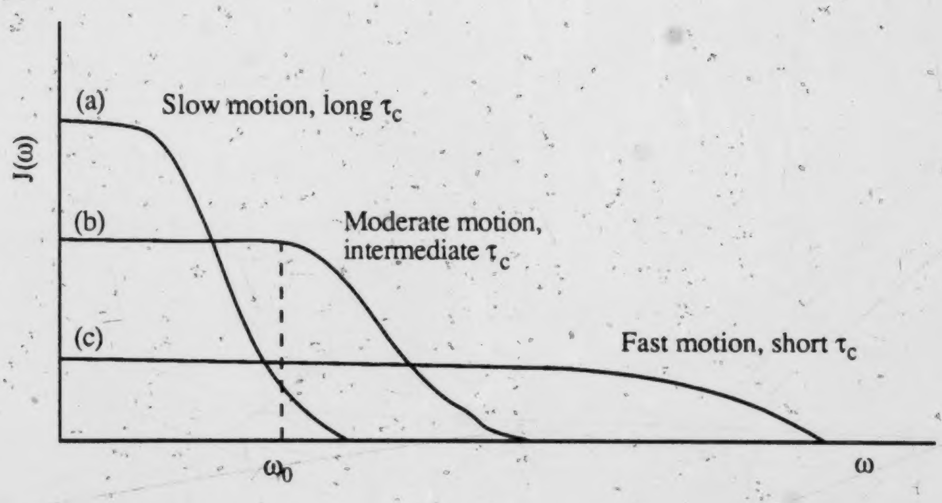


Figure 2.5.1.1 *Motional frequency spectrum at (a) slow motion, (b) moderate motion, and (c) fast motion. J(ω) represents the relative intensity at the motional frequency ω. The observing frequency for spin transitions is* ω_0.

Slow rotational motions in solids have long τ_c values, which then result in a low-frequency power spectrum of the fluctuating fields. The resulting small component at the Larmor frequency leads to a very long T_1 value. Similarly, very fast motions have short τ_c values, and hence a wide power spectrum with a small Larmor frequency component, and thus a long T_1 value. However, intermediate thermal motions of water protons bound to macromolecules in tissue, for example, result in a large Larmor frequency component and hence a short T_1 value. For solids, T_1 has been found to vary from 10^{-4} to 10^4 seconds, while for pure liquids the range is 10^{-3} to 10 seconds, but the presence of paramagnetic species may reduce the lower end of this range to 10^{-5} seconds (Emsley, 1965).

2.5.2 Transverse relaxation

The second relaxation process, called *transverse relaxation* or *spin-spin relaxation*, causes the disappearance of the transverse component of the magnetization.

The transverse component of the magnetization disappears due to the dephasing of its numerous component μ's, without any energy transfer to the lattice. If the magnetization were tipped into the xy plane with a 90° pulse, the individual spins would remain in phase with each other if each one felt exactly the same magnetic field B_0. However, in addition to interacting with the lattice, magnetic nuclei can also interact amongst themselves. These interactions of spin dipoles with other neighboring dipoles lead to the superimposition of a time dependent local magnetic field upon B_0. Suppose the net field over the sample varies over the range $B_0 \pm \Delta B$, then the set of spins which feel the field B_0 will remain fixed in the rotating frame (which rotates in the laboratory frame with a frequency $\omega_0 = \gamma B_0$); however, those spins that feel a field $B_0 + \Delta B$ will precess faster by $\Delta\omega = \gamma \Delta B$, and those experiencing a field $B_0 - \Delta B$ will precess more slowly. The net transverse component of the magnetization **M** will disappear as the spins come out of phase with each other, and the free induction signal will therefore decay, as illustrated in Figure 2.5.2.1.

The equilibrium value of M_{xy} is zero, as there is no preferred orientation within the xy plane, and therefore one way to describe the return of the magnetization to equilibrium may be

$$\frac{d}{dt} M_{xy} = -\frac{M_{xy}}{T_2} \qquad (2.5.2.1)$$

This equation expresses the exponential decay of M_{xy} to zero with a time constant T_2, called the spin-spin relaxation time (or transverse relaxation time).

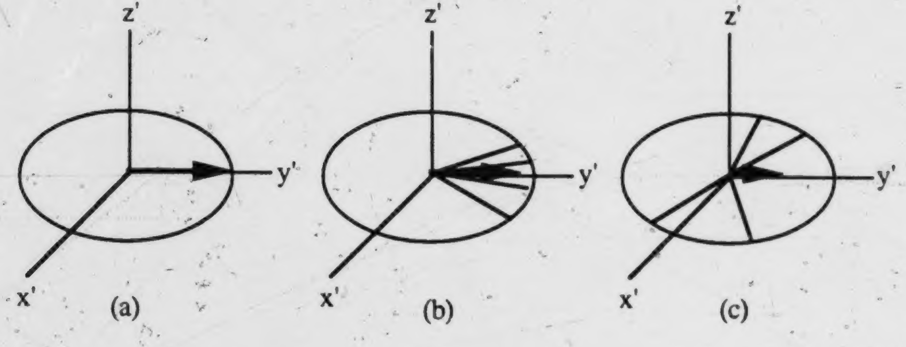


Figure 2.5.2.1 *Dephasing of spins in the rotating frame due to differences in the local magnetic field felt by the various spins. (a) represents the situation immediately after the application of a 90° pulse; the net magnetization is represented by the heavy arrow. The individual nuclear magnets gradually fan out, as shown by the light lines in (b), and (c). Therefore, the net transverse component of the magnetization gradually declines.*

In general, $T_2 \leq T_1$, with $T_2 \approx T_1$ occuring in the case of very rapid molecular motion in non-viscous liquids. In a solid or viscous liquid where the internal fields can have a large static component, $T_2 << T_1$. For pure water, $T_2 \approx T_1 \approx 3$ seconds. If additional factors such as a magnetic field inhomogeneity appear, the decay is even more rapid. The observed decay under these circumstances is usually characterized by the time constant T_2^* where $T_2^* \leq T_2$.

2.6 Bloch equations in the rotating frame

Summarizing the results of sections 2.3 and 2.5, we have the Bloch equations in the laboratory frame, which may be written in vector notation as

$$\dot{\mathbf{M}}(t) = \gamma \mathbf{M}(t) \times \mathbf{B}(t) - \boldsymbol{T}_{12} \{ \mathbf{M}(t) - \mathbf{M}_0 \}, \tag{2.6.1}$$

where $\dot{\mathbf{M}}(t) \equiv \frac{d}{dt}\mathbf{M}(t)$.

The magnetization vector **M** has the thermal equilibrium value $\mathbf{M}_0$, and $\boldsymbol{T}_{12}$ is the relaxation matrix

$$\boldsymbol{T}_{12} = \begin{bmatrix} \frac{1}{T_2} & 0 & 0 \\ 0 & \frac{1}{T_2} & 0 \\ 0 & 0 & \frac{1}{T_1} \end{bmatrix}, \tag{2.6.2}$$

with the longitudinal and transverse relaxation times T_1 and T_2, respectively. The external magnetic field **B**(t) is composed of a static field $\mathbf{B}_0$ and an RF field $\mathbf{B}_1(t)$, i.e.,

$$\mathbf{B}(t) = \mathbf{B}_0 + \mathbf{B}_1(t), \tag{2.6.3}$$

with

$$\mathbf{B}_1(t) = B_1(\cos(\omega_{rf}t + \xi)\mathbf{e}_x - \sin(\omega_{rf}t + \xi)\mathbf{e}_y). \tag{2.6.4}$$

ξ in Equation (2.6.4) is a fixed angle between the direction of the resultant $\mathbf{B}_1$ field and the x-axis (excursion from the x-axis towards y-axis). As mentioned in section 2.3.2, it is usually more convenient to calculate the response of a nuclear spin system to RF pulses in a frame rotating with the applied radio-frequency. So let us denote $\boldsymbol{R}_z(\theta)$ as the rotation matrix that will rotate the coordinate axes counterclockwise θ degrees about the z-axis, i.e.,

$$\boldsymbol{R}_z(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.6.5}$$

Assume that the pulse length is short compared to T_1 or T_2, so that the relaxation during the pulse is negligible. The magnetization vector in the rotating frame is given by

$$\mathbf{M}^r(t) = \boldsymbol{R}_z(-\omega_{rf}t)\,\mathbf{M}(t) \,, \tag{2.6.6}$$

and the Bloch equation in the rotating frame becomes

$$\dot{\mathbf{M}}^r(t) = \gamma\,\mathbf{M}^r(t) \times \mathbf{B}^r - \boldsymbol{T}_{12}\,\{\,\mathbf{M}^r(t) - \mathbf{M}_0\,\}, \tag{2.6.7}$$

where the effective magnetic field in the rotating frame $\mathbf{B}^r$ has the components

$$\begin{aligned} B_x^r &= B_1 \cos\xi \,, \\ B_y^r &= -B_1 \sin\xi \,, \\ B_z^r &= B_0 - \frac{\omega_{rf}}{\gamma} = \frac{\omega_0 - \omega_{rf}}{\gamma} = \frac{\Omega}{\gamma} \,. \end{aligned} \tag{2.6.8}$$

In explicit notation, Equation (2.6.7) is equivalent to

$$\begin{aligned} \dot{M}_x^r(t) &= \gamma\,[\,M_y^r(t)B_z^r - M_z^r(t)B_y^r\,] - \frac{1}{T_2}\,M_x^r(t) \\ \dot{M}_y^r(t) &= \gamma\,[\,M_z^r(t)B_x^r - M_x^r(t)B_z^r\,] - \frac{1}{T_2}\,M_y^r(t) \\ \dot{M}_z^r(t) &= \gamma\,[\,M_x^r(t)B_y^r - M_y^r(t)B_x^r\,] - \frac{1}{T_1}\,(\,M_z^r(t) - M_0\,) \,. \end{aligned} \tag{2.6.9}$$

2.6.1 Ideal pulse experiment

From Equation (2.6.8), if $|B_1| \gg \left|\frac{\Omega}{\gamma}\right| = \left|\frac{\omega_0 - \omega_{rf}}{\gamma}\right|$, the effective magnetic field in the rotating frame is basically in the transverse plane. An RF pulse of duration

τ_p will rotate the equilibrium magnetization M_0 about the direction of the applied RF field B_1, irrespective of the resonance offset Ω, by (Equation (2.3.2.4))

$$\beta = \gamma B_1 \tau_p . \qquad (2.6.1.1)$$

For an RF field applied along the x-axis ($\xi = 0$ in Equation (2.6.4)), the initial magnetization just after the pulse will be

$$\begin{aligned} M_x^r(0_+) &= 0 , \\ M_y^r(0_+) &= M_0 \sin\beta , \\ M_z^r(0_+) &= M_0 \cos\beta . \end{aligned} \qquad (2.6.1.2)$$

The subsequent magnetization in the rotating frame can be described as follows :

$$\begin{aligned} M_x^r(t) &= M_0 \sin\beta \sin(\Omega t) \exp(-t/T_2) , \\ M_y^r(t) &= M_0 \sin\beta \cos(\Omega t) \exp(-t/T_2) , \\ M_z^r(t) &= M_0 + (M_0 \cos\beta - M_0) \exp(-t/T_1) ; \end{aligned} \qquad (2.6.1.3)$$

or in the laboratory frame,

$$\begin{aligned} M_x(t) &= M_0 \sin\beta \sin(\omega_0 t) \exp(-t/T_2) , \\ M_y(t) &= M_0 \sin\beta \cos(\omega_0 t) \exp(-t/T_2) , \\ M_z(t) &= M_0 + (M_0 \cos\beta - M_0) \exp(-t/T_1) . \end{aligned} \qquad (2.6.1.4)$$

Hence, after the RF pulse the transverse magnetization will be precessing about the axis of B_0 with angular frequency ω_0; but in contrast to section 2.3.1, it is decaying exponentially with the time constant T_2. At the same time, M_z recovers with the time constant T_1.

2.6.2 The effect of finite RF pulse amplitude

In the previous section, we assumed that the amplitude of the applied RF pulse was strong enough to allow the neglect of all off-resonance effects. However, if nuclei are present in several different chemical environments, the resulting transverse magnetization will contain contributions with several different precessional frequencies. As a result of this, the width of the spectrum to be investigated may be comparable to the maximum available RF field strength B_1, and off-resonance effects would not under these circumstances be negligible. We now consider this off-resonance effect.

As shown in Figure 2.6.2.1, the tilted effective field about which the magnetization rotates is determined by the offset field

$$\Delta B_0 = B_0 - \frac{\omega_{rf}}{\gamma} = \frac{\Omega}{\gamma} \tag{2.6.2.1}$$

along the z-axis and by the RF field B_1 in the transverse plane. It has the amplitude

$$B_{eff} = \sqrt{B_1{}^2 + (\Delta B_0)^2} \;, \tag{2.6.2.2}$$

and is tilted by an angle θ with respect to the z-axis, where

$$\tan\theta = \frac{B_1}{\Delta B_0} = \frac{\gamma B_1}{\Omega} \;. \tag{2.6.2.3}$$

The effective nutation angle β of the magnetization vector about $\mathbf{B}_{\mathbf{eff}}$ in the clockwise direction during a pulse of duration τ_p is

$$\beta = \gamma B_{eff}\, \tau_p = \gamma B_1 \tau_p \sqrt{1 + \left(\frac{\Omega}{\gamma B_1}\right)^2} \;. \tag{2.6.2.4}$$

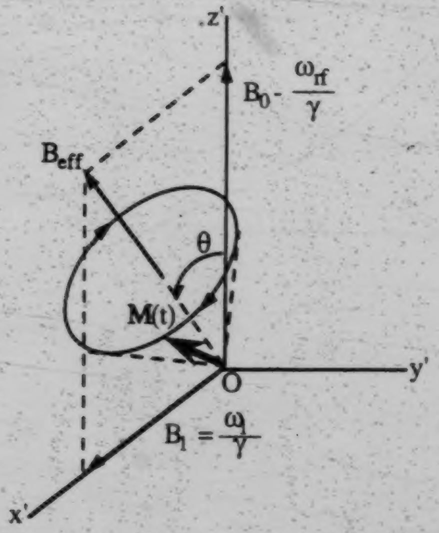

z'
$B_0 - \frac{\omega_{rf}}{\gamma}$
B_{eff}
θ
M(t)
O
y'
$B_1 = \frac{\omega_1}{\gamma}$
x'

Figure 2.6.2.1 Tilted effective field in the rotating frame. θ is defined in Equation (2.6.2.3).

For an RF field applied along the x'-axis of the rotating frame, i.e., $\xi = 0$ in Equation (2.6.4), the components of the magnetization immediately after the RF pulse can be computed as follows :

$$\mathbf{M}^r(0_+) = \boldsymbol{R}_{y'}^{-1}(\theta)\ \boldsymbol{R}_{z'}(\beta)\ \boldsymbol{R}_{y'}(\theta)\ \mathbf{M}^r(0_-)\,, \qquad (2.6.2.5)$$

with

$$\boldsymbol{R}_{y'}(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \qquad (2.6.2.6)$$

where $\boldsymbol{R}_{z'}$ is defined as in Equation (2.6.5), and $\mathbf{M}^r(0_-)$ represents the magnetization in the rotating frame just before the RF pulse.

If $\mathbf{M}^r(0_-) = M_0\, \mathbf{e}_{z'}$, Equation (2.6.2.5) results in

$$\begin{aligned} M_x^r(0_+) &= M_0\,(1 - \cos\beta)\sin\theta\cos\theta\,, \\ M_y^r(0_+) &= M_0 \sin\beta \sin\theta\,, \\ M_z^r(0_+) &= M_0\,(\cos^2\theta + \cos\beta\sin^2\theta)\,. \end{aligned} \qquad (2.6.2.7)$$

The transverse magnetization immediately after the RF pulse is no longer along the y'-axis, as in the case of on-resonance irradiation, but has developed a phase shift φ, measuring from x'-axis towards y'-axis, which depends on the resonance offset Ω as well as the effective pulse rotation angle β_{eff},

$$\begin{aligned} \tan\varphi &= \frac{M_y^r(0_+)}{M_x^r(0_+)} \\ &= \frac{\sin\beta}{(1-\cos\beta)} \times \frac{1}{\cos\theta} \\ &= \frac{\sin\beta}{(1-\cos\beta)} \times \frac{\sqrt{1+\left(\frac{\Omega}{\gamma B_1}\right)^2}}{\frac{\Omega}{\gamma B_1}}\,. \end{aligned} \qquad (2.6.2.8)$$

For the on resonance condition and a strong B_1 field applied along the x'-axis, we have $\xi = 0$ and $B_{eff} = B_1$, and hence, $\theta = 90°$. Substituting these results in Equation (2.6.2.8), we have $\varphi = 90°$. This means that the magnetization just after the RF pulse is always in the y'z'-plane, which is consistent with Equation (2.6.1.2).

At the same time, the amplitude response of the transverse magnetization is given by

$$M^r{}_y(0_+) = \sqrt{[M_x^r(0_+)]^2 + [M_y^r(0_+)]^2}$$

$$= M_0 \sin\theta \sqrt{(1-\cos\beta)^2\cos^2\theta + \sin^2\beta}. \qquad (2.6.2.9)$$

Figure 2.6.2.2 shows the dependence of the signal phase φ and effective pulse rotation angle β as functions of $\Omega/(\gamma B_1)$. Figure 2.6.2.3 shows the amplitude response of the transverse magnetization versus $\Omega/(\gamma B_1)$. Note that the response remains almost constant up to an offset Ω approximately equal to the RF field strength γB_1. However, for larger offsets, the response decreases to zero, then increases again and starts to oscillate as a function of the offset. Vanishing transverse magnetization implies that the effective pulse rotation angle β_{eff} about the tilted axis is a multiple of 2π, and the magnetization vector is rotated back along the z-axis at the end of the RF pulse. Some typical ratios of spectral width and B_1 field strength in *in-vivo* spectroscopy are shown in Table 2.6.2.1 for $B_0 = 2.35$ T and B_1 in the range of 10^{-5} T to 10^{-3} T. For the definition of the unit 'ppm', please see Appendix 3. This range of B_1 field corresponds to pulse length τ_p from about 500 µs to 5 µs for effective pulse rotation angle $\beta_{eff} = 90°$ for protons.

Nucleus	$\gamma \times 10^8$ $(\frac{rad}{Ts})$	υ_0 (MHz)	γB_1 (kHz)	Approximate chemical shift, Ω (ppm)	$\frac{\Omega}{\gamma B_1}$
^1H	2.6752	100	0.4258 - 42.58	10	0.0235 - 2.349
^{13}C	0.6726	25.156	0.1070 - 10.70	200	0.4702 - 47.02
^{31}P	1.0829	40.502	0.1723 - 17.23	40	0.094 - 9.403

Table 2.6.2.1 *Approximate spectral width and B_1 field strength in* in-vivo *spectroscopy, assuming $B_0 = 2.35$ T and B_1 is in the range of 10^{-5} T to 10^{-3} T.*

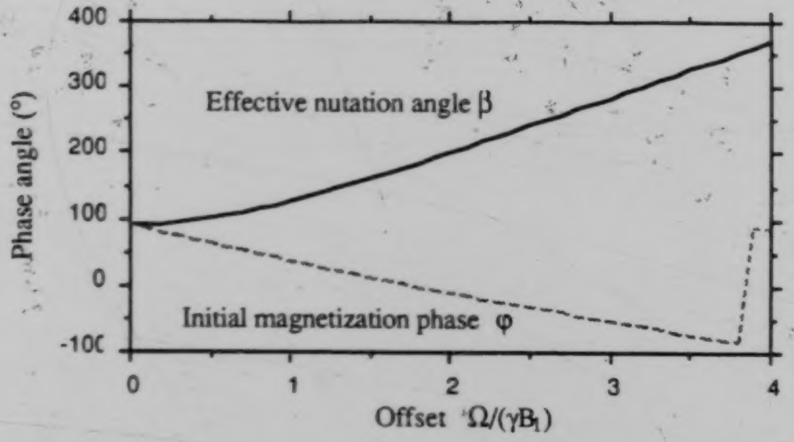


Figure 2.6.2.2 *Effect of offset $\Omega/(\gamma B_1)$ on the signal phase defined in Equation (2.6.2.8) and effective pulse rotation angle β defined in Equation (2.6.2.4) for a single-pulse experiment. The pulse width τ_p is assumed to be a constant such that $\beta = 90°$ at $\Omega = 0$ throughout all frequency offsets.*

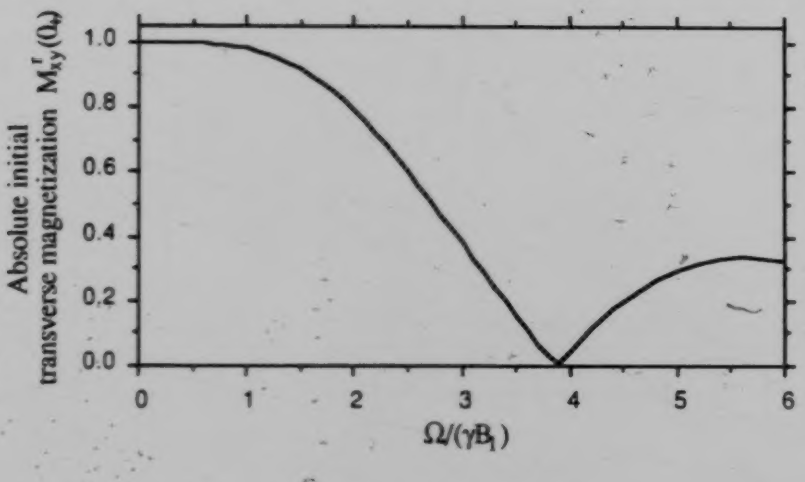


Figure 2.6.2.3 *Effect of offset $\Omega/(\gamma B_1)$ on transverse magnetization $M^r_{xy}(0_+)$ defined in Equation (2.6.2.9) for a single-pulse experiment. The pulse width τ_p is assumed to be a constant such that $\beta = 90°$ at $\Omega = 0$.*

If $\mathbf{M}^r(0_-) = M_0\,\mathbf{e}_z$, and the RF pulse is applied along an x'-axis which is ξ degree away from the x-axis to the y-axis at time t = 0 as shown in Equation (2.6.4), then one can show that the subsequent magnetization in the rotating frame can be described by

$$\begin{aligned} M_x^r(t) &= M_{xy}^r(0_+)\sin(\Omega t-\varphi)\exp(-t/T_2)\ , \\ M_y^r(t) &= M_{xy}^r(0_+)\cos(\Omega t-\varphi)\exp(-t/T_2)\ , \\ M_z^r(t) &= M_0 + (M_z^r(0_+)-M_0)\exp(-t/T_1)\ , \end{aligned} \tag{2.6.2.10}$$

or equivalently, in the laboratory frame,

$$\begin{aligned} M_x(t) &= M_{xy}^r(0_+)\sin(\omega_0 t+\xi-\varphi)\exp(-t/T_2)\ , \\ M_y(t) &= M_{xy}^r(0_+)\cos(\omega_0 t+\xi-\varphi)\exp(-t/T_2)\ , \\ M_z(t) &= M_0 + (M_z^r(0_+)-M_0)\exp(-t/T_1)\ . \end{aligned} \tag{2.6.2.11}$$

Equation (2.6.2.11) is expanded from

$$\mathbf{M}(t) = \boldsymbol{R}_z(\omega_{rf}t+\xi)\,\mathbf{M}^r(t)\ . \tag{2.6.2.12}$$

Notice that φ is a function of the resonance offset $\Omega = \omega_0 - \omega_{rf}$.

2.7 Detection of NMR signals

Let us now consider the detection of the NMR signal. In the last few sections, we have shown that by applying an RF field with the appropriate radio-frequency and pulse width, one can obtain a net transverse magnetization. This transverse magnetization will precess about the direction of B_0 and decay due to factors like transverse relaxation and field inhomogeneity after the RF pulse is terminated. If a conducting coil is placed so that its axis lies in the transverse plane, there will be a time dependent change of magnetic flux through this coil as the magnetization precesses. Faraday's law of electromagnetic induction says that the induced electromotive force (EMF) in a circuit is proportional to the rate at which the magnetic flux through a closed circuit is changing. Therefore, an EMF will be induced in the coil by the rotating magnetization. This is the origin of the measured NMR signal.

Suppose we have nuclei present in K different environments; then from Equation (2.6.2.11), and Faraday's law of induction, we have

$$\mathrm{EMF} \propto \sum_{i=1}^{K} M_{xy}^{i}(0_{+}) \cos(\omega_i t + \varsigma_i)\, e^{-t/T_{2i}} , \qquad (2.7.1)$$

where

$M_{xy}^{i}(0_{+})$ = transverse magnetization just after the RF pulse from i-th environment in the laboratory frame;

ω_i = angular frequency of the transverse magnetization from i-th environment ;

ς_i = sum of the phase of the receiving coil with the initial transverse magnetization, ξ, and the phase problem due to factors like off-resonance effect, φ_i;

T_{2i} = decay constant of the transverse magnetization from i-th environment .

This decaying sinusoidal signal is commonly called the *free induction decay (FID)* signal. It describes the decay of the induced signal arising from free precession of the nuclei in the field $\mathbf{B}_0$.

Usually the NMR signal detected from the receiver coil is in the form of a very small voltage, typically a few 10's of μV's for *in-vitro* samples or 10's of mV's for *in-*

vivo samples. It is therefore essential to preamplify the signal before any further processing. Figure 2.7.1.1 shows a simplified block diagram of the receiver chain of a typical pulsed NMR spectrometer.

Since the frequencies in Equation (2.7.1) are typically on the order of tens of megahertz and occupy a total spectral width of only a few kilohertz, it is difficult for the computer to sample at such high frequencies. Moreover, the filter has difficulty selecting a region occupying only a few kilohertz at a frequency of tens of megahertz. Therefore, the approach taken is to subtract the frequency $\omega_{rf}/(2\pi)$ of the applied B_1 field from the frequencies of the signals, thereby generating a group of frequencies in the range zero to a few kilohertz. These frequencies, which are in the audio-frequency range, can be handled by the computer and filtered adequately. The required frequency translation is accomplished with the aid of a device called a phase-sensitive detector. This device multiplies the incoming signal by a sine-wave of some arbitrary reference frequency. In NMR the reference frequency is usually derived from the same RF generator used to provide the excitation of the spin system. Assume that the incoming signal to the phase-sensitive detector, neglecting the proportionality constant, is given by Equation (2.7.1),

$$s(t) = \sum_{i=1}^{K} M_{xy}^{i}(0_+) \cos(\omega_i t + \varsigma_i)\, e^{-t/T_{2i}} , \tag{2.7.2}$$

then the output $v_1(t)$ from the detector is

$$\begin{aligned} v_1(t) &= s(t) \cos(\omega_{rf} t) \\ &= \sum_{i=1}^{K} M_{xy}^{i}(0_+) \cos(\omega_i t + \varsigma_i) \cos(\omega_{rf} t)\, e^{-t/T_{2i}} \end{aligned}$$

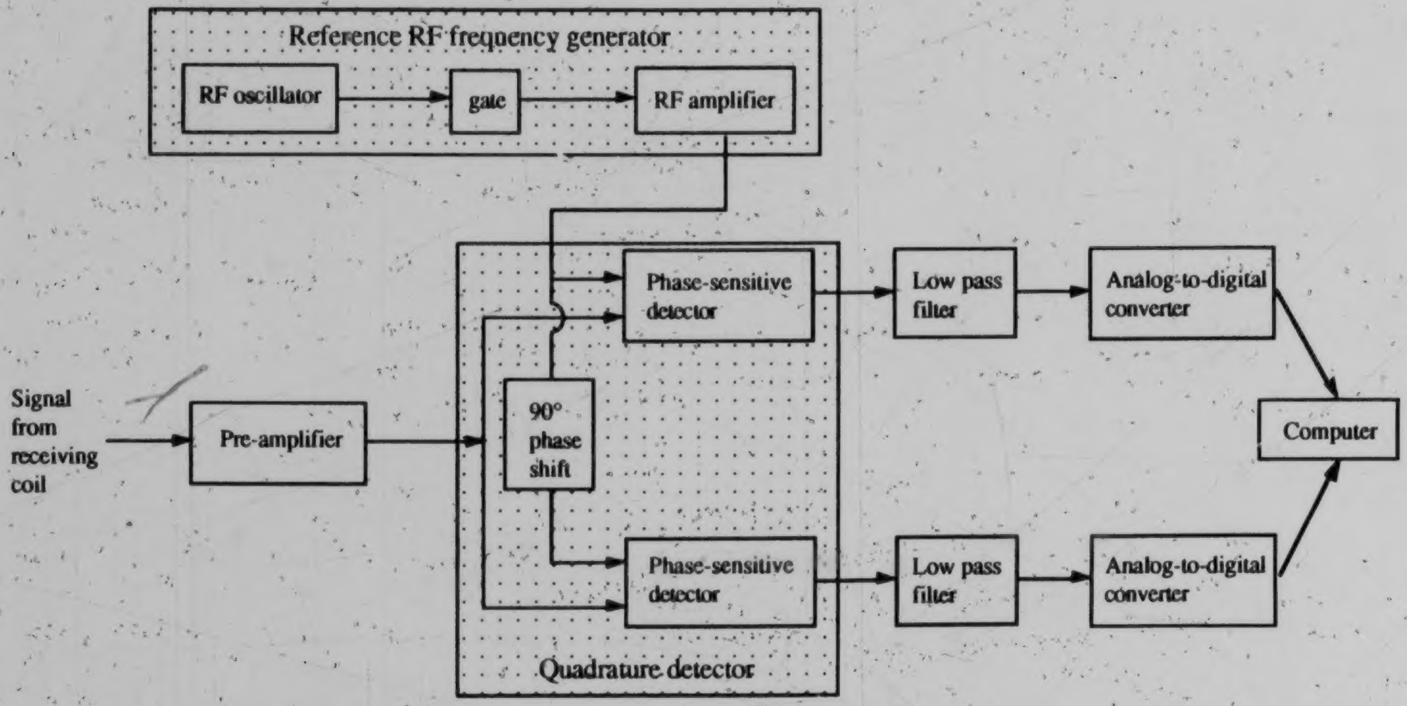

Reference RF frequency generator
RF oscillator
gate
RF amplifier
Signal from receiving coil
Pre-amplifier
90° phase shift
Phase-sensitive detector
Phase-sensitive detector
Quadrature detector
Low pass filter
Low pass filter
Analog-to-digital converter
Analog-to-digital converter
Computer

Figure 2.7.1.1 *A simplified block diagram of pulsed NMR spectrometer*

$$v_1(t) = \frac{1}{2}\sum_{i=1}^{K} M_{xy}^{i}(0_+) \cos[(\omega_i - \omega_{rf})t + \varsigma_i]\, e^{-t/T_{2i}} + \frac{1}{2}\sum_{i=1}^{K} M_{xy}^{i}(0_+) \cos[(\omega_i + \omega_{rf})t + \varsigma_i]\, e^{-t/T_{2i}} \ .$$

Thus, the output of the phase-sensitive detector consists of two types of sine wave, one with angular frequency of $(\omega_i - \omega_{rf})$ and the other with angular frequency $(\omega_i + \omega_{rf})$. This output is then filtered by a low-pass filter of bandwidth 2W, resulting in the passage of only those components having angular frequencies of $(\omega_i - \omega_{rf})$ to pass through, i.e.,

$$v_2(t) = \frac{1}{2}\sum_{i=1}^{K} M_{xy}^{i}(0_+) \cos(\Omega_i t + \varsigma_i)\, e^{-t/T_{2i}} \ , \qquad (2.7.3)$$

where $\Omega_i = \omega_i - \omega_{rf}$.

This output is converted to a digital signal by an analog-to-digital converter, for which the sampling rate is at least W/π.

If we have only one channel of signal available for Fourier transformation, all signals will appear on one side of the reference frequency, regardless of the sign of $\omega_i - \omega_{rf}$. To overcome this problem, we usually use two phase-sensitive detectors, as illustrated in Figure 2.7.1, but the phase of the reference is shifted by 90° in one of the detectors. The outputs from the two phase-sensitive detectors differ in phase by 90°. This combination is called quadrature detection. The outputs from the quadrature detector are both filtered and sampled before being fed into a computer. The result may be represented by

$$v_3(t) = \frac{1}{2}\sum_{i=1}^{K} M_{xy}^{i}(0_+) \cos(\Omega_i t + \varsigma_i)\, e^{-t/T_{2i}} + j\frac{1}{2}\sum_{i=1}^{K} M_{xy}^{i}(0_+) \sin(\Omega_i t + \varsigma_i)\, e^{-t/T_{2i}} \ ,$$

$$= \sum_{i=1}^{K} A_i \, e^{j\varsigma_i} \, e^{(-1/T_{2i}+j\Omega_i)t} \quad . \qquad (2.7.4)$$

Therefore, the outputs from the filters after the quadrature detector are effectively the components of the transverse magnetization in the rotating frame, each of which are basically a sum of exponentially decaying sinusoidal signals. After analog-to-digital conversion, the recorded FID may then be represented by Equation (1.1.1).

2.8 Conventional signal processing in pulsed NMR spectroscopy

Following accumulation of the FID, the resulting signal must be transformed in order to produce a spectrum in which signal amplitude is plotted as a function of frequency. This is usually done by Fourier transformation. In addition to this process, a number of additional steps are also required, and in this section we discuss briefly all the important aspects of data processing.

2.8.1 Fourier transformation

In traditional pulsed NMR spectroscopy, the parameters that we want to estimate, namely the number of exponentially decaying components, the amplitudes, the phases, the damping rates and the frequencies of each component in the FID, are usually estimated from the spectrum which is obtained by discrete Fourier transformation (DFT) of the sampled FID. Before we define the relationship of Fourier transformation between the discrete-time signal $s(n\Delta t)$ and the discrete-frequency function $S(mF)$, let us define the following quantities :

t = continuous-time variable;

Δt = time interval between successive samples when a time function is digitized, also called dwell time ($\Delta t = 1 / f_s$);

t_p = effective period of data collection, also called acquisition time ($t_p = 1/F = N\Delta t$);

f = continuous-frequency variable;

F = frequency increment between successive components when a frequency function is sampled ($F = 1 / t_p$);

f_s = sampling rate when a time function is sampled, i.e., the number of samples per second ($f_s = 1 / \Delta t = NF$);

N = number of samples in the range $0 \leq t < t_p$ when the time function is sampled. N is also equal to the number of samples in the range $0 \leq f < f_s$ when the frequency function is sampled.

Some of the above quantities are illustrated in Figure 2.8.1.1. The discrete Fourier transform pair of $s(n\Delta t)$ and $S(mF)$ is defined as follows :

$$S(mF) = \Delta t \sum_{n=0}^{N-1} s(n\Delta t) \exp(-j2\pi \frac{mn}{N})$$

$$s(n\Delta t) = \frac{1}{N\Delta t} \sum_{m=0}^{N-1} S(mF) \exp(j2\pi \frac{mn}{N}) . \qquad (2.8.1.1)$$

If the noise-free time-domain signal is described as in Equation (1.1.1),

$$s(n\Delta t) = A\, e^{j\phi} e^{2\pi(-\alpha + jf)n\Delta t}, \qquad n = 0, \ldots, N-1, \quad (2.8.1.2)$$

where α is positive, then one can show that the first order approximation of its discrete Fourier transform is equivalent to the continuous Fourier transform of the continuous-time signal provided $\alpha N\Delta t$ is much less than zero, and also that aliasing, leakage and picket-fence effects are negligible (one can find more about these effects from Stanley et al., 1984, p.283). The first order approximation of the discrete Fourier transform of Equation (2.8.1.2) is as follows :

$$S(mF) \approx \frac{A\,e^{j\phi}}{2\pi\alpha - j2\pi(f - \frac{m}{N\Delta t})}$$

$$= \frac{A\,\alpha\cos\phi - A\,(f - \frac{m}{N\Delta t})\sin\phi}{2\pi\alpha^2 + 2\pi(f - \frac{m}{N\Delta t})^2} + j\,\frac{A\,\alpha\sin\phi + A\,(f - \frac{m}{N\Delta t})\cos\phi}{2\pi\alpha^2 + 2\pi(f - \frac{m}{N\Delta t})^2}\,. \tag{2.8.1.3}$$

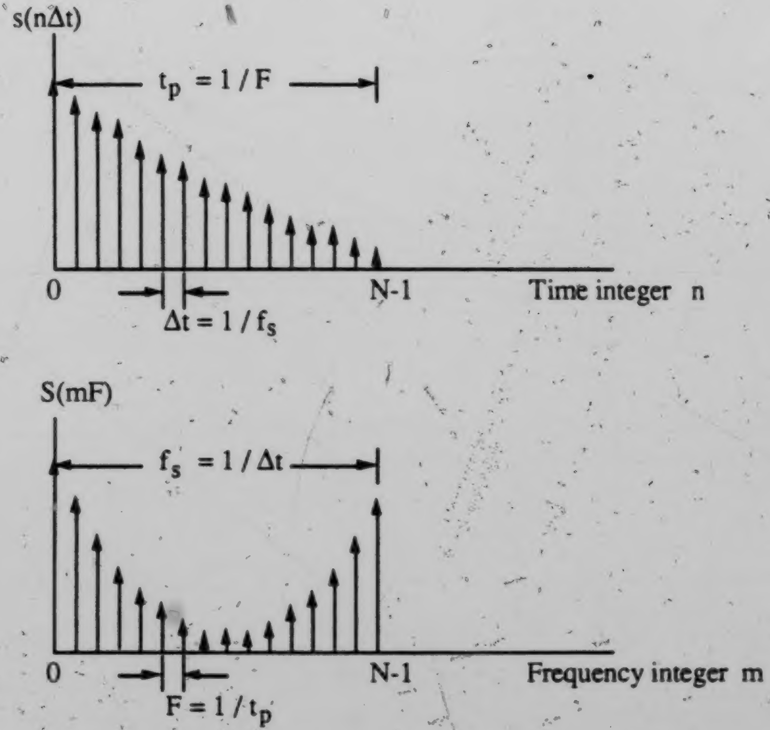

s(nΔt)
t_p = 1 / F
0
N-1
Time integer n
Δt = 1 / f_s
S(mF)
f_s = 1 / Δt
0
N-1
Frequency integer m
F = 1 / t_p

Figure 2.8.1.1 Illustration of a Fourier transform pair

The aliasing effect can be minimized by sampling the data at at least twice the frequency of the signal; leakage and picket-fence effects can be minimized by increasing the number of data points being sampled.

If the spectrum is phase corrected, the real part of it has the characteristic Lorentzian lineshape of the form given by

$$g(mF) = \frac{A\alpha}{2\pi\alpha^2 + 2\pi(f - \frac{m}{N\Delta t})^2} \quad . \qquad (2.8.1.4)$$

This is commonly termed the absorption mode, and it is usually this part of the spectrum that is finally displayed.

2.8.2 Zero filling

Zero filling involves adding an array of zeros to the end of the FID. One reason for doing this is that the typical fast Fourier transform algorithm requires the number of data points be 2 to the power of some integer. So if the sampled FID does not meet this criterion, then zero filling is needed. The second reason is that by padding zeros to the end of the FID, the density of points in the transformed spectrum is increased. By sufficient zero padding of the given N samples before the DFT, we can locate the frequency of an isolated spectral peak as accurately as we wish. However, the spectral resolution does not change because the spectral resolution that can be achieved by zero padding is limited by the bandwidth of the rectangular window transform, which is about $2\pi/(N\Delta t)$, and the linewidth of the peaks.

2.8.3 Baseline correction

Sometimes an NMR spectrum exhibits a baseline distortion. One possible cause of this problem is that there are spike signals at the beginning of the FID.

Traditionally, there are at least two ways to handle this problem. Firstly, replace the first few FID data points by zeros. An example of this is illustrated in Figure 2.8.3.1 where the first FID has strong spike signals at the beginning. This FID then has subtracted from it a second signal which has only the spike signal in the first part and the rest of it is zeros. The resulting signal is shown in Figure 2.8.3.1(c). If we let the first FID be $s_1(n)$ and the second signal be $s_2(n)$, then by the linearity property of the DFT, the above operation is equivalent to subtracting the DFT of $s_1(n)$ by the DFT of $s_2(n)$. Therefore, this is a reasonable operation provided the region of spike signals is very short compared with the data acquistation period, i.e., just a few sampled points. Practically, this solution leads to another problem, and that is, the difficulty in justifying the region of spike signals. If one deletes too many data points, one destroys information about the correct signals, especially for fast decaying signals. In addition, after baseline correction, one usually needs to phase correct the spectrum. If it is not possible to phase correct it, then one needs to go back and try another baseline correction.

A second way of handling this baseline problem is by fitting arbitrary functions to the spectrum in place of the spikey region around $t = 0$.

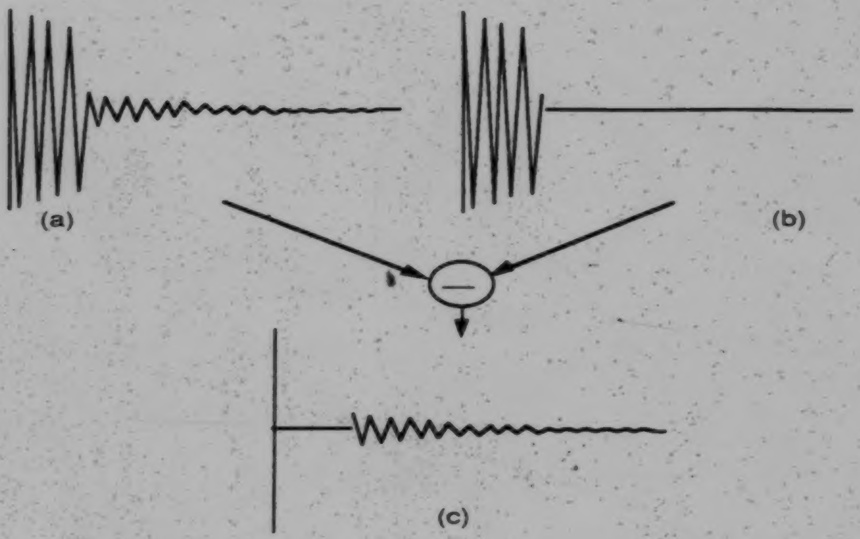


Figure 2.8.3.1 Illustration of one method of baseline correction

2.8.4 Phase correction

The aim of phase correction is to ensure that all the signals in the real part of the transformed spectrum are in the absorption mode. Phase problems can be caused by many things. For instance,

(i) finite delay between the end of the RF pulse and the start of data acquisition;
(ii) finite RF pulse amplitude as mentioned in section 2.6.2;
(iii) phase difference between the RF field and the receiver, like the angle ξ defined in Equation (2.6.4);
(iv) baseline correction, etc.

A typical way to alleviate this problem is to use a first order phase correction given by

$$\phi(f_p,f) = P_0 + P_1 (f - f_p) \quad . \qquad (2.8.4.1)$$

where f_p is called the pivot point. One first selects the pivot point, then adjusts P_0 and P_1 until the final spectrum is phase corrected. If the spectrum cannot be phase corrected with all possible combinations of P_0 and P_1, then one needs to choose another pivot point and go through the whole process again. Sometimes this can be an extremely difficult process, especially if there is a baseline problem in the spectrum. This practice assumes that the phase of the whole spectrum is linear with respect to a pivot point frequency. If that is not the case, then people usually correct a smaller region of the spectrum and ignore the rest of it.

2.8.5 Apodization

In practical Fourier spectroscopy, the acquisition time t_p for the FID is always limited and the signal s(t) may have a sharp cut-off at the end. The truncated signal can be thought of as the product of the untruncated signal with a rectangular weighting function. The corresponding Fourier spectrum is then obtained by convolution of the undistorted spectrum with the Fourier transform of the rectangular weighting function. This Fourier transform of the rectangular weighting function produces oscillating signal tails (ripples or sidelobes). The peak sidelobe of a rectangular window is about -13 dB from the mainlobe level and the sidelobes thereafter fall off by only 6 dB/octave[1]. This severely limits the ability of the DFT to resolve a weak signal in the proximity of a stronger one. It is the purpose of apodization to modify the envelope of the truncated signal by multiplication with a weighting function such that these sidelobes are largely suppressed. Typical apodization (window) functions are :

---

[1] For window amplitude response function in dB scale, see Appendix 5.

1. mono-exponential window
2. cosine window
3. Hanning window
4. Hamming window
5. Blackman window
6. Kaiser window

The price we pay for the improvements of peak sidelobe and sidelobe rolloff rate is a reduced spectral resolution because these windowing function have a larger mainlobe width. For example, the mainlobe width of the Hanning window is approximately twice that of the rectangular window transform, but the peak sidelobe is approximately -31 dB and the sidelobe rolloff rate is about 18 dB/octave for the Hanning window. The other problem associated with apodization is that the information about the original signals might be altered by the operation. One can find more about apodization functions in Ernst et al. (1988, p.101-106) or Jackson (1989, p.145-152).

2.8.6 Resolution enhancement

Resolution enhancement attempts to achieve a transformation of the lineshape to narrow the resonance lines artificially. In practice, resolution enhancement can only be achieved at the expense of reducing the signal-to-noise ratio (SNR). This is because resolution enhancement necessarily implies the enhancement of the latter part of the FID signal, and as a result random noise contributions in the latter part of the signal may also be enhanced. In order to obtain a compromise between resolution and sensitivity, some people have designed resolution enhancement functions that decay towards zero at the end of the FID. Some commonly used resolution enhancement methods are :

1. multiplying the FID by a growing exponential ;
2. convolution difference ;
3. Lorentz-Gauss transformation ;

The ultimate resolution enhancement requires complete flattening of the FID by multiplication with the inverse envelope of the FID, leading to a rectangular FID of duration t_p. The lineshape will now exhibit excessive ripple. The full width at half-height of the mainlobe of the spectrum is

$$\Delta f \approx 0.604 / t_p .$$

This is the minimum achievable width. Hence, we see that the achievable resolution enhancement is restricted by the acquisition time t_p. Notice again that resolution enhancement methods often generate some distortion of the spectral lines, and great care must be taken when interpreting the resulting peak areas and linewidths.

Chapter 3

Pseudo-inversion and singular value decomposition

3.1 Introduction

In section 1.4.1, we showed that the periodogram cannot be used to estimate frequencies and damping rates of sinusoids that are not well resolved or are not close to non-decaying. Then, in section 1.4.2, we mentioned AR spectral estimation where we focused our attention on Prony's method. We showed that Prony's method tried to find least squares estimates of the g(k)'s by minimizing Equation (1.4.2.2.6) with respect to the g(k)'s. Equation (1.4.2.2.6) is written here again:

$$\begin{aligned} E &= \frac{1}{N-M} \sum_{n=M}^{N-1} | \varepsilon(n) |^2 \\ &= \frac{1}{N-M} \sum_{n=M}^{N-1} | y(n) + \sum_{k=1}^{M} g(k)\, y(n-k) |^2 . \end{aligned} \tag{3.1.1}$$

But unfortunately, Prony's method attempted to minimize the 'false' error equation (3.1.1), not the true least squares error $\sum_{n=0}^{N-1} | y(n) - \hat{y}(n) |^2$, where $\hat{y}(n)$ is the estimated signal at time n. So in general, it produces an inferior 'fit' to the data. Kumaresan (1982) tried to increase the order of the AR model employed (i.e., he increased M in Equation (3.1.1) to L where $M < L \leq N-M$). This meant that he modified the error equation to

$$\begin{aligned} E &= \frac{1}{N-L} \sum_{n=L}^{N-1} | \varepsilon(n) |^2 \\ &= \frac{1}{N-L} \sum_{n=L}^{N-1} | y(n) + \sum_{k=1}^{L} g(k)\, y(n-k) |^2 , \end{aligned} \tag{3.1.2}$$

or equivalently solved

$$\begin{bmatrix} y(L-1) & y(L-2) & \cdots & y(0) \\ y(L) & y(L-1) & \cdots & y(1) \\ \vdots & \vdots & & \vdots \\ y(N-2) & y(N-3) & \cdots & y(N-1-L) \end{bmatrix} \begin{bmatrix} g(1) \\ g(2) \\ \vdots \\ g(L) \end{bmatrix} = - \begin{bmatrix} y(L) \\ y(L+1) \\ \vdots \\ y(N-1) \end{bmatrix},$$

$$\hat{\boldsymbol{A}}\,\mathbf{g} = -\hat{\mathbf{h}}\,, \qquad (3.1.3)$$

in a least squares sense for the g(k)'s, where $k = 1, 2, \ldots, L$. The hat ^ means that the corresponding variable is affected by noise or estimated under noise. For noiseless quantities, the hat ^ is dropped. Kumaresan observed that at low SNRs, increasing the AR model order may result in spurious peaks. He then attempted to use only a subset of the L exponentials that best explained the data. This resulted in principal component estimation and noise subspace estimation methods.

Now, let us look at Equation (3.1.3) again (because this is what the linear prediction singular value decomposition algorithm will use). One can show that Equation (3.1.3) is equivalent to Equation (1.4.2.2.10) mathematically, but not necessarily numerically. By that we mean that solving Equation (1.4.2.2.10) by computer is usually numerically poorer than solving Equation (3.1.3) directly because of the round-off errors introduced by multiplications in forming the elements of $\hat{\boldsymbol{R}}_{yy} = \hat{\boldsymbol{A}}^{H}\hat{\boldsymbol{A}}$ and $\hat{\mathbf{r}}_{yy} = \hat{\boldsymbol{A}}^{H}\hat{\mathbf{h}}$ in Equation (1.4.2.2.10), and also due to problems associated with the property of the condition number of $\hat{\boldsymbol{R}}_{yy}$ and that of $\hat{\boldsymbol{A}}$ ($\kappa_2(\hat{\boldsymbol{R}}_{yy}) = \kappa_2^2(\hat{\boldsymbol{A}})$). The condition number of a matrix $\hat{\boldsymbol{A}}$, $\kappa_2(\hat{\boldsymbol{A}})$, is defined as

$$\kappa_2(\hat{\boldsymbol{A}}) = \|\hat{\boldsymbol{A}}\|_2 \, \|\hat{\boldsymbol{A}}^{+}\|_2\,, \qquad (3.1.4)$$

which tells us how errors in the given system of equations may be magnified in the least squares solution (Stewart, 1973, p. 221-230). $\|\hat{\boldsymbol{A}}\|_2$ is the 2-norm of $\hat{\boldsymbol{A}}$, and $\hat{\boldsymbol{A}}^{+}$ is the pseudo-inverse of $\hat{\boldsymbol{A}}$. Therefore, if it is possible, one prefers solving Equation (3.1.3) to Equation (1.4.2.2.10).

The least squares solution of **g** in Equation (3.1.3) is

$$\hat{\mathbf{g}} = -(\hat{A}^H\hat{A})^{-1}\hat{A}^H\hat{\mathbf{h}} . \qquad (3.1.5)$$

Equation (3.1.5) assumes that $\hat{A}$ is full column rank so that the inverse of $\hat{A}^H\hat{A}$ exists. Unfortunately, one can show that in the noiseless case A is not full column rank, but of rank M only, so that the inverse of $A^H A$ does not exist! To offset this, we introduce another concept, namely, the *pseudo-inverse* of a general matrix.

3.2 Pseudo-inversion and minimum 2-norm least squares solution

It is well known that if A is a square non-singular matrix, then there exists a matrix G which is called the inverse of A and denoted by A^{-1}, such that $AG = GA = I$, where I is the identity matrix. If A is singular or rectangular, no such matrix G exists. Now, if A^{-1} exists, then the system of linear equations $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$. However, in many cases solutions of a system of linear equations exist even when the inverse of the matrix defining these equations does not. For example, if L = M and there is no noise in the linear system defined in Equation (3.1.3), the matrix A^{-1} does not exist because A is rectangular; but the solution **g** exists, as shown in Equation (1.4.2.2.2). In cases where there is no solution to a system of linear equations, we call these equations inconsistent (in other words, there is no **x** such that the equality $A\mathbf{x} = \mathbf{b}$ holds). These problems, along with others in numerical linear algebra, control, optimization, statistics, and other areas of analysis, are readily handled via the concept of a generalized inverse of a matrix.

In 1920, Moore (1920) extended the notion of the "inverse" of a matrix to singular or rectangular matrices. His definition of the generalized inverse of an m x n matrix $\boldsymbol{A}$, is equivalent to the existence of an n x m matrix, $\boldsymbol{G}$ such that

$$\boldsymbol{A}\,\boldsymbol{G} = \boldsymbol{P}_A, \qquad \boldsymbol{G}\,\boldsymbol{A} = \boldsymbol{P}_G,$$

where $\boldsymbol{P}_A$ stands for the projection operator onto the range of $\boldsymbol{A}$, i.e., the space generated by the columns of $\boldsymbol{A}$ (denoted by $R(\boldsymbol{A})$). In 1955, Penrose (1955) rediscovered Moore's general inverse, which he called the "generalized inverse". Penrose's definition of the generalized inverse is as follows: for any matrix $\boldsymbol{A}$, square or rectangular, there exists a unique matrix $\boldsymbol{G}$ satisfying the four conditions :

$$\boldsymbol{A}\,\boldsymbol{G}\,\boldsymbol{A} = \boldsymbol{A} \tag{3.2.1}$$

$$\boldsymbol{G}\,\boldsymbol{A}\,\boldsymbol{G} = \boldsymbol{G} \tag{3.2.2}$$

$$(\boldsymbol{AG})^{\mathrm{H}} = \boldsymbol{AG} \tag{3.2.3}$$

$$(\boldsymbol{GA})^{\mathrm{H}} = \boldsymbol{GA} \tag{3.2.4}$$

where "H" denotes complex conjugate transposition. These conditions are equivalent to Moore's conditions. The unique matrix $\boldsymbol{G}$ that satisfies these relations is now known as the "Moore-Penrose pseudo-inverse", and is denoted by $\boldsymbol{A}^{+}$.[1]

The applications of pseudo-inverse matrices are diverse, e.g., least squares, linear equations, projections, statistical regression analysis, filtering, linear programming, optimization, nonlinear analysis, control and system identification, games theory, network and econometrics. A deeper understanding of these topics is achieved when they are studied in the generalized inverse context (Pringle and Rayner,1971; Albert,1972; Klema and Laub, 1980). Handy bibliographies on pseudo-inverse matrices can be found in Boullion and Odell (1968), and Nashed (1976).

---

[1] Some writers have adopted descriptive names to designate various classes of generalized inverses. But there is a lack of consistency in the use of these terms by different writers. A $\boldsymbol{G}$ which satisfies (3.2.1) is called *a generalized inverse* (Rao, 1965; Boullion and Odell, 1968), *pseudo-inverse* (Sheffield, 1958), *inverse* (Bjerhammar, 1973), or *g_1-inverse* (Pringle and Rayner, 1971). A $\boldsymbol{G}$ which satisfies (3.2.1) and (3.2.2) is called a *semi-inverse* (Frame, 1964), *reciprocal inverse* (Bjerhammar, 1973), *reflexive generalized inverse* (Rohde, 1965; Boullion and Odell, 1968), or *g_2-inverse* (Pringle and Rayner, 1971). A $\boldsymbol{G}$ which satisfies (3.2.1-3.2.3) is called a *normalized generalized inverse* (Rohde, 1965), *left weak generalized inverse* (Boullion and Odell, 1968), or *g_3-inverse* (Pringle and Rayner, 1971). A $\boldsymbol{G}$ which satisfies (3.2.1), (3.2.2) and (3.2.4) is called a *right weak generalized inverse* (Boullion and Odell, 1968), or *g^{*}_3-inverse* (Pringle and Rayner, 1971). A $\boldsymbol{G}$ which satisfies (3.2.1-3.2.4) is called *the general reciprocal* (Moore, 1920), *generalized inverse* (Penrose), *pseudoinverse* (Greville, 1959; Boullion and Odell, 1968), the *Moore-Penrose inverse* (Rao and Mitra, 1971), *Moore-Penrose pseudoinverse* (Albert, 1972), or *the g-inverse* (Pringle and Rayner, 1971). In view of this diversity of terminology, the *pseudo-inverse* is used to represent a matrix that satisfies (3.2.1-3.2.4) in the whole thesis.

In this thesis, we will use pseudo-inverse matrices to obtain least squares solutions. But before we proceed to do so, we state the following theorems. The proofs of these theorems are given in Appendices 6 and 7.

Theorem 3.2.1 : The pseudo-inverse $\boldsymbol{A}^{+}$ of an m x n complex matrix $\boldsymbol{A}$ is unique.

Theorem 3.2.2 : The matrix $G = (\boldsymbol{A}^{H}\boldsymbol{A})^{+}\boldsymbol{A}^{H} = \boldsymbol{A}^{H}(\boldsymbol{A}\boldsymbol{A}^{H})^{+}$ is the pseudo-inverse of $\boldsymbol{A}$.

The uniqueness of the pseudo-inverse of a given matrix is important because if there were many pseudo-inverses for one given matrix, pseudo-inverse matrices might not be as useful. As a consequence of this uniqueness property, we need not worry about getting many pseudo-inverses from a given matrix because there is only one that exists and it always exists. Theorem 3.2.2 gives us two equalities for the pseudo-inverse of $\boldsymbol{A}$. Note that this definition applies to any kind of matrix: non-singular, singular, rectangular, etc. One can use either one of the equalities, whichever is more convenient to the problem. For example, if the columns of $\boldsymbol{A}$ are linearly independent, so that $\boldsymbol{A}^{H}\boldsymbol{A}$ is non-singular, then $(\boldsymbol{A}^{H}\boldsymbol{A})^{+} = (\boldsymbol{A}^{H}\boldsymbol{A})^{-1}$, and one would choose $(\boldsymbol{A}^{H}\boldsymbol{A})^{+}\boldsymbol{A}^{H}$ as the pseudo-inverse of $\boldsymbol{A}$. On the other hand, if the rows of $\boldsymbol{A}$ are linearly independent, $\boldsymbol{A}\boldsymbol{A}^{H}$ is non-singular, then $(\boldsymbol{A}\boldsymbol{A}^{H})^{+} = (\boldsymbol{A}\boldsymbol{A}^{H})^{-1}$, and hence one would choose $\boldsymbol{A}^{H}(\boldsymbol{A}\boldsymbol{A}^{H})^{+}$ as the pseudo-inverse of the matrix $\boldsymbol{A}$.

As mentioned above, the pseudo-inverse of a general matrix is very useful in many disciplines. One particularly important application of the pseudo-inverse matrix applicable to us is that it can be used to obtain a least squares solution of a system of equations, and this least squares solution has one important property. Let us review some properties of solutions of a system of equations, then we will show that important property associated with the pseudo-inverse.

For a given m x n complex matrix $\boldsymbol{A}$ (denoted by $\boldsymbol{A} \in C^{m \times n}$) and an m-dimensional complex vector b (denoted by $\mathbf{b} \in C^m$), the linear system

$$\boldsymbol{A}\mathbf{x} = \mathbf{b} \tag{3.2.5}$$

is consistent (i.e., the system has a solution of $\mathbf{x}$) if and only if b is in the range of $\boldsymbol{A}$ (this means that b is a linear combination of the vectors in $\boldsymbol{A}$, and is denoted by $\mathbf{b} \in R(\boldsymbol{A})$). Otherwise, the residual vector

$$\mathbf{r} = \mathbf{b} - \boldsymbol{A}\mathbf{x} \tag{3.2.6}$$

is nonzero for all $\mathbf{x} \in C^n$. In this case it may be desired to find an *approximate* solution of Equation (3.2.5), by which is meant a vector **x** making the residual vector defined by Equation (3.2.6) "closest" to zero in some sense, in other words, minimizing a norm of Equation (3.2.6). An approximation that is often used is the *least squares* solution, defined as the vector **x** which minimizes the sum of the squares of the moduli of the residuals **r**. For convenience, let us denote $\| \mathbf{x} \|_2$ as the square root of the sum of the squares of the moduli of the elements in **x**, i.e.,

$$\| \mathbf{x} \|_2 = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{1/2} = (\mathbf{x}^H \mathbf{x})^{1/2}, \tag{3.2.7}$$

where **x** is a complex n-vector (denoted by $\mathbf{x} \in C^n$), and "H" denotes complex conjugate transposition. Equation (3.2.7) is often called the 2-norm (or Euclidean norm) of vector $\mathbf{x}$.

From now on, we will focus our attention on the case where the linear system in Equation (3.2.5) is inconsistent, and we want to find a least squares solution of the problem, i.e., we want to minimize the function

$$\| \mathbf{r} \|_2^2 \;=\; \| \mathbf{b} - \boldsymbol{A}\mathbf{x} \|_2^2 \,, \tag{3.2.8}$$

with respect to $\mathbf{x}$. The following theorem tells us the existence of least squares solution and its uniqueness. Its proof can be found in Stewart (1973, p. 220).

Theorem 3.2.3 : The linear least squares problem of minimizing Equation (3.2.8) with respect to $\mathbf{x}$ always has a solution which is given by

$$\mathbf{x}_0 \;=\; \boldsymbol{A}^{+}\mathbf{b} \;+\; (\boldsymbol{I} - \boldsymbol{R}_A)\mathbf{z} \,, \tag{3.2.9}$$

where $\boldsymbol{R}_A = \boldsymbol{A}^{+}\boldsymbol{A}$ is the projection onto the row space of $\boldsymbol{A}$, and $\mathbf{z}$ is arbitrary. The solution is unique if and only if the nullity of $\boldsymbol{A}$ (denoted by null($\boldsymbol{A}$)) is equal to zero (that means, rank($\boldsymbol{A}$) = number of columns in $\boldsymbol{A}$).

Going back to our problem, i.e., finding the least squares solution of Equation (3.1.3), one can show that in the noiseless case and for $M < L \leq N\text{-}M$, null($\boldsymbol{A}$) is always greater than zero because $\boldsymbol{A}$ has L-M linearly dependent columns. Therefore, from Theorem 3.2.3, the least squares solution of Equation (3.1.3) is not unique.

Of all the possible least squares solutions $\hat{\mathbf{g}}$'s, one of them is extremely useful in the development of the LPSVD algorithm. This is the one which has the minimum sum of squares. We denote this particular solution as $\hat{\mathbf{g}}_{LS}$. In the following theorem, we will show that $\hat{\mathbf{g}}_{LS} = -\boldsymbol{A}^{+}\mathbf{h}$ is *either* the unique vector (approximate solution or exact solution) for which $\| \boldsymbol{A}\hat{\mathbf{g}} + \mathbf{h} \|_2$ is minimized, *or* is the unique vector for which both $\| \boldsymbol{A}\hat{\mathbf{g}} + \mathbf{h} \|_2$ and $\| \hat{\mathbf{g}} \|_2$ is minimized. The proof is given in Appendix 8.

Theorem 3.2.4 : Let $A \in C^{m \times n}$, $b \in C^m$ and $x_{LS} = A^+ b$. Then each $x \in C^n$, $x \neq x_{LS}$, satisfies one of the two conditions :

(a) $\| A x_{LS} - b \|_2 < \| A x - b \|_2$ or

(b) $\| A x_{LS} - b \|_2 = \| A x - b \|_2$ and $\| x_{LS} \|_2 < \| x \|_2$.

Theorem 3.2.4 can be restated as follows: let $A \in C^{m \times n}$ and $b \in C^m$, then amongst the least squares solutions of $Ax = b$, A^+b is the one of minimum 2-norm. Conversely, if $G \in C^{n \times m}$ has the property that, for all b, $G b$ is the minimum 2-norm least squares solution of $Ax = b$, then $G = A^+$. This is the property of pseudo-inversion that will be used in the development of the LPSVD algorithm (see Theorem 4.2.1.3).

3.3 Singular value decomposition (SVD)

There are a few ways to calculate the pseudo-inverse of a general matrix. Some of them are described in Albert (1972, chapter 5), Golub and Van Loan (1985, chapter 6). Among all of these techniques, singular value decomposition is one of the most powerful, especially in solving the rank deficient least squares problem[1]. The existence of SVD of a matrix is established in the following theorem.

Theorem 3.3.1 : Singular value decomposition (SVD)

Let A be an m x n matrix of rank r and, without lost of generality, assume that m>n. Then, there exists two unitary matrices U of order m and V of order n such that

---

[1] In fact, Golub and Van Loan (1985, p. 170) said that SVD is the only reliable way to treat rank deficient least squares problem.

$$U^{H} A V = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \tag{3.3.1}$$

where Σ_r is an r x r diagonal matrix which contains the square root of the nonzero eigenvalues σ_i^2's of the matrix AA^H arranged in the order of $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_r^2$.

Proof :

Since AA^H is positive semi-definite, its eigenvalues are non-negative. Let them be $\sigma_1^2, \sigma_2^2, \ldots\ldots, \sigma_n^2$, where $\sigma_1^2 \geq \sigma_2^2 \geq \ldots\ldots \geq \sigma_r^2 > 0 = \sigma_{r+1}^2 = \sigma_{r+2}^2 = \ldots\ldots = \sigma_n^2$. Let U be a set of orthonormal eigenvectors of AA^H corresponding to its eigenvalues. Then we have

$$A A^H U = U \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Let $U = [\, U_1, U_2 \,]$, then

$$U^H A A^H U = \begin{bmatrix} U_1^H A A^H U_1 & U_1^H A A^H U_2 \\ U_2^H A A^H U_1 & U_2^H A A^H U_2 \end{bmatrix} = \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Consequently, we have

$$U_1^H A A^H U_1 = \Sigma_r^2 \quad \Rightarrow \quad \Sigma_r^{-1} U_1^H A A^H U_1 \Sigma_r^{-1} = I, \tag{3.3.2}$$

$$U_2^H A A^H U_2 = 0 \quad \Rightarrow \quad A^H U_2 = 0.$$

Let $V_1 = A^H U_1 \Sigma_r^{-1}$. Then from Equation (3.3.2) $V_1^H V_1 = I$, that is, the columns of V_1 are orthonormal. Let V_2 be chosen so that $V = [V_1, V_2]$ is unitary. Then

$$U^H A V = \begin{bmatrix} U_1^H A V_1 & U_1^H A V_2 \\ U_2^H A V_1 & U_2^H A V_2 \end{bmatrix}$$

$$= \begin{bmatrix} U_1^H A (A^H U_1 \Sigma_r^{-1}) & (V_1 \Sigma_r^{-1})^H V_2 \\ 0(V_1) & 0(V_2) \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}.$$

The σ_i's are called the *singular values* of A and the column vectors of U and V are, respectively, called the *left singular vectors* (the orthonormal eigenvectors of AA^H) and the *right singular vectors* (the orthonormal eigenvectors of $A^H A$) of A.

To see the relationship of the pseudo-inverse and SVD of a matrix, one can substitute Equation (3.3.1) into Theorem 3.2.2 as follows:

$$A^+ = A^H (A A^H)^+$$

$$= V \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} U^H \left[U \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix} U^H \right]^+$$

$$= V \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} U^H U \begin{bmatrix} \Sigma_r^{-2} & 0 \\ 0 & 0 \end{bmatrix} U^H$$

$$= V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^H . \qquad (3.3.3)$$

An alternate proof of the relationship of pseudo-inverse and SVD can be found in Pringle and Rayner (1971, p. 4).

The SVD has very important applications to the linear least squares problem. This can be seen by substituting Equation (3.3.3) into Theorem 3.2.4. We have the following theorem which tells us how to obtain the minimal 2-norm least squares solution of $\boldsymbol{A}\mathbf{x} = \mathbf{b}$ from SVD.

Theorem 3.3.2 : If $\boldsymbol{A}^{+}$ is the pseudo-inverse of $\boldsymbol{A} \in \mathrm{C}^{m \times n}$ having rank $r \leq \min(m,n)$, then the minimal 2-norm least squares solution of $\boldsymbol{A}\,\mathbf{x} = \mathbf{b}$ is

$$\mathbf{x}_{LS} = \boldsymbol{A}^{+}\mathbf{b} = \sum_{i=1}^{r} \left(\frac{\mathbf{u}_i^H \mathbf{b}}{\sigma_i}\right)\mathbf{v}_i \ , \qquad (3.3.4)$$

and the minimum residual sum of squares is

$$r_{LS}^2 = \| \boldsymbol{A}\,\mathbf{x}_{LS} - \mathbf{b} \|_2^2 = \sum_{i=r+1}^{\max(m,n)} \left(\mathbf{u}_i^H \mathbf{b}\right)^2 \ , \qquad (3.3.5)$$

where the $\mathbf{u}_i$'s and $\mathbf{v}_i$'s are the column vectors of $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively.

Hence, by using the above theorem, the minimum 2-norm least squares solution of Equation (3.1.3) is

$$\mathbf{g}_{LS} = -\boldsymbol{A}^{+}\mathbf{h} = -\sum_{i=1}^{M} \left(\frac{\mathbf{u}_i^H \mathbf{h}}{\sigma_i}\right)\mathbf{v}_i \ , \qquad (3.3.6)$$

where the $\mathbf{u}_i$'s and $\mathbf{v}_i$'s are the column vectors of $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively, and

$$\boldsymbol{A}^{+} = \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Sigma}_M^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{U}^{H} \ ,$$

and rank($\boldsymbol{A}$) = $M \leq p = \min(N-L, L)$. As mentioned before, this particular solution will be used in the development of the LPSVD algorithm.

3.4 SVD and rank determination

The above introduction of pseudo-inversion is mainly based on finding the inverse of a rank deficient matrix because rank(***A***) defined in Equation (3.1.3) is equal to the number of exponential components in an FID (i.e., M). In theory, this presents no problem. One applies the SVD on $\boldsymbol{A} \in C^{N-L \times L}$ and observes that if rank(***A***) = M, where $0 < M < p = \min(N-L,L)$, then

$$\sigma_{M+1}^{2} = \sigma_{M+2}^{2} = \ldots\ldots = \sigma_{p}^{2} = 0 .$$

In practice, rounding errors and noise are present in the system. The given noise corrupted matrix $\hat{\boldsymbol{A}}$ in Equation (3.1.3) is no longer defective in rank, but is near a matrix of defective rank. Standard methods of solving Equation (1.4.2.2.10) may take this matrix and turn it into one that is clearly of full rank. This introduces large errors in the least squares solution which may explain why Kumaresan (1982) observed that at low SNRs, increasing the AR model order resulted in spurious peaks. Also, recall from section 1.4.3 that from the theoretical autocorrelation point of view, it makes sense to retain only the M principal eigenvector components in the estimate of the autocorrelation matrix because the effect of noise is to introduce noise eigenvectors with, hopefully, small eigenvalues into the signal autocorrelation matrix. Therefore, it is necessary to obtain an estimate of the rank of matrix $\hat{\boldsymbol{A}}$ in Equation (3.1.3). In fact, rounding errors and fuzzy data make rank determination a nontrival exercise because they perturb the singular values. For example, suppose we have an FID composed of three components and sampled at 2500 Hz. The parameters of the components are listed in the following tabulation :

amplitude	phase [rad/sec]	normalized damping rate	normalized frequency[1]
250	0	0.003183	-0.3
200	0	0.012732	-0.28
200	0	0.013369	-0.25

The effect of noise on the singular values of $\hat{\boldsymbol{A}}$ given by Equation (3.1.3) for the case L = 75 and N = 150 is demonstrated in the following table and Figure 3.4.2. Representative spectra are shown in Figure 3.4.1. From Figure 3.4.2, one finds that in the noiseless case, σ_3 is far greater than σ_4, σ_5 and so on, which indicates that the rank of $\boldsymbol{A}$ is 3. This is exactly equal to the number of components in the FID as one would predict. But as the noise level increases, σ_3 becomes less well separated from the other smaller singular values. Increasing the noise variance further will eventually make even σ_2 hard to distinguish from the rest of the smaller singular values.

singular values	time domain noise standard deviation σ_W			
	0	15	30	45
σ_1	5888.60	5887.09 ± 3.869	5892.94 ± 7.719	5906.3 ± 11.535
σ_2	1238.06	1262.30 ± 1.642	1334.16 ± 3.240	1457.58 ± 4.577
σ_3	473.54	521.18 ± 2.215	778.44 ± 2.959	1147.90 ± 4.260
σ_4	0.00	381.38 ± 1.456	714.96 ± 2.093	1054.04 ± 3.102
σ_5	0.00	349.28 ± 1.067	674.45 ± 1.801	996.57 ± 2.639

***Table 3.4.1** Demonstration of the effect of noise on the singular values of $\hat{\boldsymbol{A}}$ given by Equation (3.1.3).*

---

[1] The relationship of normalized frequency and frequency in units of ppm is shown in Appendix 3. For example, if the spectrometer frequency is 100 MHz, then the frequencies of the three components in units of ppm will be -7.5, -7.0 and -06.25. On the other hand, if the spectrometer frequency is 40.55 MHz, then the frequencies become -18.496, -17.263 and -15.413 ppm.

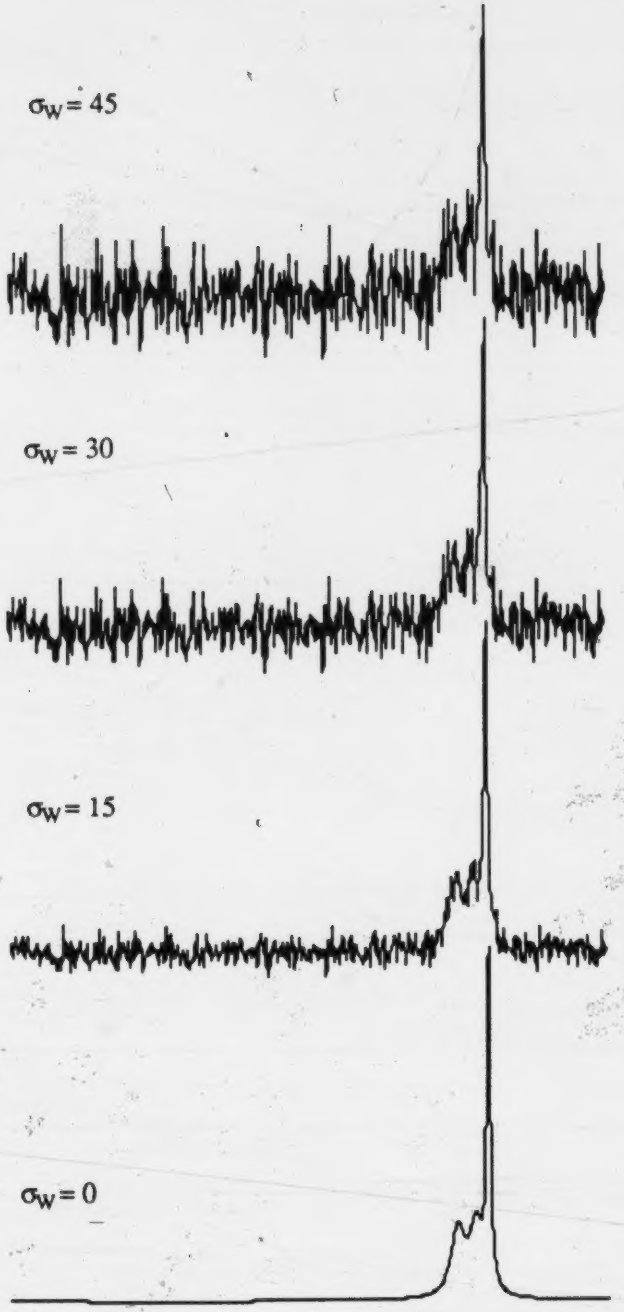


Figure 3.4.1 *The range of simulated spectra used to demonstrate the effect of noise on the singular values.*

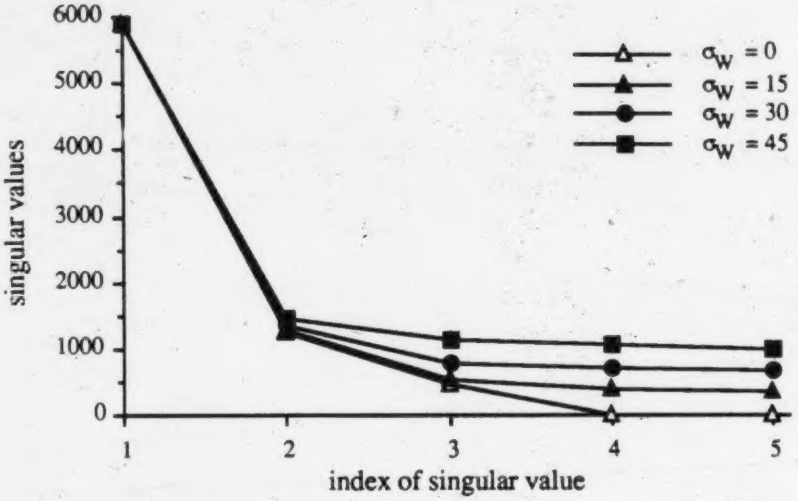


Figure 3.4.2 Graphical presentation of the results in Table 3.4.1.

To demonstrate the combined effects of near rank deficiency and noise on the singular values of $\hat{A}$, consider the following example. Suppose now we have an FID consisting of three components and again sampled at 2500 Hz. The parameters of the components are listed as follows :

amplitude	phase [rad/sec]	normalized damping rate	normalized frequency
250	0	0.003183	-0.3
200	0	œ	-0.28
200	0	0.013369	-0.25

where the œ was varied in three simulation studies as shown below :

simulation	œ
1	0.003183
2	0.009549
3	0.015915

We first observed the singular values of $\hat{A}$ given by Equation (3.1.3), where L was set to 75 and N to 150 in the noiseless case. Then, sequences of Gaussian noise of variance $2\sigma_W^2 = 2(15)^2 = 450$ were added to the FIDs. Six hundred independent tries were performed in each simulation. Representative spectra are shown in Figure 3.4.3. Table 3.4.2 and Figure 3.4.4 show the largest five singular values of $\hat{A}$ in each simulation. From Figure 3.4.4, one makes the following observations. Firstly, increasing damping rate of an exponential component in an FID can make the matrix $\hat{A}$ in Equation (3.1.3) closer to rank deficiency. Secondly, the same sets of noise can affect the rank of a rank deficient matrix more seriously than that of a rank non-deficient matrix. From both Figure 3.4.2 and Figure 3.4.4, one notices that the principal singular values are not perturbed as seriously as the extraneous singular values.

σ_W	singular values	normalized damping rate œ		
		0.003183	0.009549	0.015915
0	σ_1	5810.39	5833.21	5928.88
	σ_2	4208.98	1390.08	1213.20
	σ_3	961.69	668.65	339.10
	σ_4	0.01	0.00	0.00
	σ_5	0.01	0.00	0.00
15	σ_1	5810.48 ± 3.522	5831.67 ± 3.865	5927.40 ± 3.873
	σ_2	4214.95 ± 3.654	1410.03 ± 1.986	1238.51 ± 1.678
	σ_3	991.58 ± 2.171	703.18 ± 2.309	422.24 ± 1.660
	σ_4	382.69 ± 1.499	382.27 ± 1.497	371.90 ± 1.205
	σ_5	349.99 ± 1.072	349.60 ± 1.071	346.16 ± 0.989

***Table 3.4.2** Influence of near rank deficiency and noise on the singular values of $\hat{A}$.*

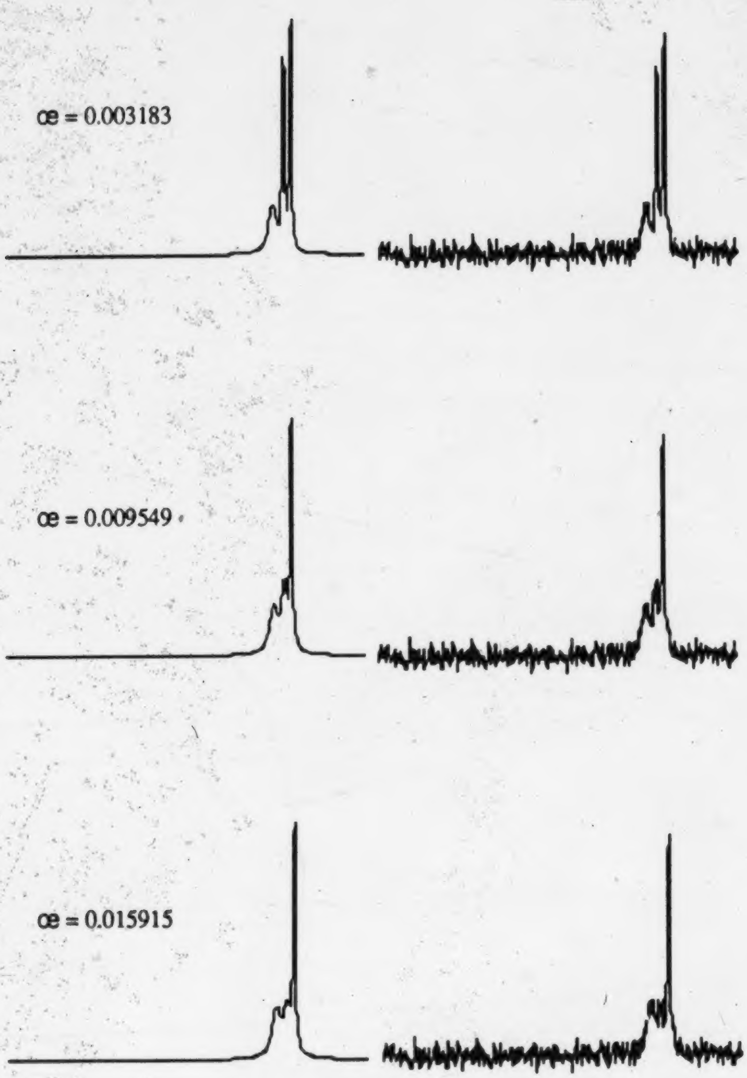


Figure 3.4.3 *The range of simulated spectra used to demonstrate the effect of near rank deficiency and noise on the singular values : (left) noiseless spectra, (right) noise corrupted spectra.*

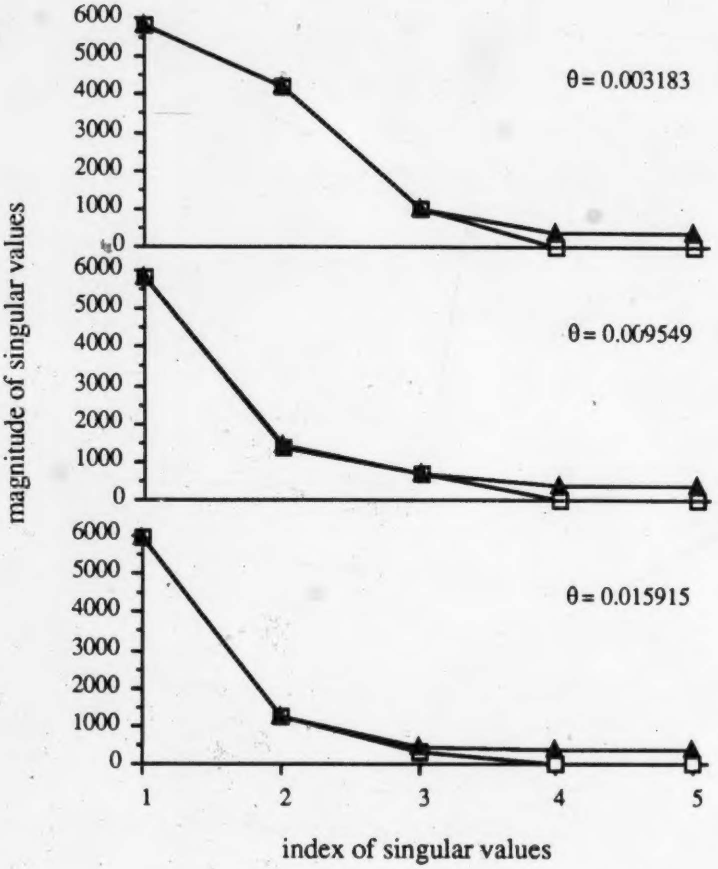


***Figure** 3.4.4* *Graphical presentation of the results in Table 3.4.2.*
—□— *no noise;* —▲— *with noise of* $\sigma_W = 15$.

From the above examples, one sees that the mere observation of small singular values does not solve the ill-conditioned least squares problem because we must decide upon a value for rank($\hat{\boldsymbol{A}}$). One approach to this difficult problem is to have a parameter $\delta > 0$ and a convention that $\hat{\boldsymbol{A}}$ has "numerical rank" $\hat{M}$ if $\hat{\sigma}_{\hat{M}}$ satisfies

$$\hat{\sigma}_1 \geq \ldots \geq \hat{\sigma}_{\hat{M}} > \delta \geq \hat{\sigma}_{\hat{M}+1} \geq \ldots \geq \hat{\sigma}_p .$$

When this is the case, we can regard

$$\hat{\mathbf{g}}_{\hat{M}} = -\hat{\boldsymbol{A}}^{+}_{\hat{M}} \hat{\mathbf{h}} = -\sum_{i=1}^{\hat{M}} \left(\frac{\hat{\mathbf{u}}_i^{H} \hat{\mathbf{h}}}{\hat{\sigma}_i} \right) \hat{\mathbf{v}}_i , \tag{3.4.1}$$

as an approximation to $\hat{\mathbf{g}}_{LS}$. Since $\| \hat{\mathbf{g}}_{\hat{M}} \|_2 \approx 1/\hat{\sigma}_{\hat{M}} \leq 1/\delta$, δ may be chosen with the intention of producing an approximate least squares solution with suitably small norm.

If $\hat{\sigma}_{\hat{M}} >> \delta$, then one may be comfortable with $\hat{\mathbf{g}}_{\hat{M}}$ because $\hat{\boldsymbol{A}}$ can then be unambiguously regarded as a rank $\hat{M}$ matrix. Yet $\{\hat{\sigma}_1, \ldots, \hat{\sigma}_p\}$ might not clearly split into subsets of large and small singular values, making the determination of $\hat{M}$ by this mean somewhat arbitrary. So far not much has been done in analyzing this difficult problem, but we will do some simulation studies on the problem where signal singular values and noise singular values are indistinguishable later in chapter 5.

Chapter 4

Two algorithms

4.1 Introduction

In section 2.6, we demonstrated that theoretically, digitized FIDs may be represented by multi-exponentially decaying sinusoids as in Equation (1.1.1). But in section 1.4.1, we showed that the periodogram cannot be used to estimate frequencies and damping rates of sinusoids that are not well resolved or are not close to non-decaying. Then, in section 1.4.2, we discussed AR spectral estimation, where we focused our attention on Prony's method. But unfortunately, Prony's method results in poor estimates for noisy data. Kumaresan and Tufts (1982) attempted to increase the order of the AR model employed (i.e., increased M in Equation (3.1.1) to L where $M < L \leq N-M$) and applied singular value decomposition (SVD) to filter the data. This, hopefully, will result in a promising method.

In the following section, we will briefly describe the Kumaresan-Tufts (KT) method, also called the Tufts-Kumaresan (TK) or the linear prediction singular value decomposition (LPSVD) method. Then, in section 4.2.1, we will show how the LPSVD method behaves in the noiseless situation. In section 4.2.2, we will show how it behaves in a noisy case. In section 4.3, we will give a proposed algorithm.

4.2 LPSVD method

Step 1 : Form the backward linear prediction data matrix $\hat{A}_B$ and data vector $\hat{h}_B$, respectively, as follows :

$$\hat{A}_B = \begin{bmatrix} y^*(1) & y^*(2) & \cdots & y^*(L) \\ y^*(2) & y^*(3) & \cdots & y^*(L+1) \\ \vdots & \vdots & & \vdots \\ y^*(N-L) & y^*(N-L+1) & \cdots & y^*(N-1) \end{bmatrix} \quad (4.2.1)$$

$$\hat{h}_B = [\, y^*(0), y^*(1), \ldots\ldots, y^*(N-L-1)\,]^T , \quad (4.2.2)$$

where "*" denotes complex conjugation; "T" denotes transposition; and L should satisfy $M \leq L \leq N-M$; and the y(i)'s are defined in Equation (1.1.1).

Step 2 : Form the coefficient vector $\hat{g}_B$ of a polynomial of order L by

$$\hat{g}_B = -[\hat{A}_B]^+_{\hat{M}}\, \hat{h}_B , \quad (4.2.3)$$

where $[\hat{A}_B]^+_{\hat{M}}$ is the "truncated rank $\hat{M}$" pseudo-inverse of $\hat{A}_B$, which is defined with the use of SVD as in Equation (3.4.1):

$$[\hat{A}_B]^+_{\hat{M}} = \sum_{i=1}^{\hat{M}} \frac{1}{\hat{\sigma}_i} \hat{v}_i \hat{u}_i^H , \quad (4.2.4)$$

where $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \ldots \geq \hat{\sigma}_{\hat{M}} \geq \ldots\ldots$ are the singular values of $\hat{A}_B$, and the $\hat{u}_i$'s and $\hat{v}_i$'s are the corresponding left and right singular vectors, respectively. The superscript "H" denotes complex conjugate transposition. It is clear from section 3.4 that for the noiseless case, $[\hat{A}_B]^+_{\hat{M}} = A_B^+$.

Step 3 : Find the zeros of the polynomial equation

$$1 + \sum_{k=1}^{L} \hat{g}_{B,k}\, \hat{z}^{-k} = 0 \quad (4.2.5)$$

where $\hat{g}_{B,k}$ is the k-th element of $\hat{g}_B$. The $\hat{M}'$ ($\hat{M}'$ may be different from $\hat{M}$ if SNR is low) zeros which are outside the unit circle are chosen as estimates of the zeros $z_i = \exp[2\pi(\alpha_i + jf_i)]$. From these zeros one can find the

$\exp[2\pi(-\hat{\alpha}_i + j\hat{f}_i)]$'s by reflecting them inside the unit circle. These zeros are called signal zeros.

Step 4 : Find the solution $\hat{\mathbf{a}}$ of

$$\hat{\boldsymbol{Z}}\hat{\mathbf{a}} = \hat{\mathbf{y}} \;, \tag{4.2.6}$$

$$\hat{\boldsymbol{Z}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{\hat{s}_1} & e^{\hat{s}_2} & \cdots & e^{\hat{s}_{\hat{M}'}} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ e^{\hat{s}_1(N-1)} & e^{\hat{s}_2(N-1)} & \cdots & e^{\hat{s}_{\hat{M}'}(N-1)} \end{bmatrix} , \tag{4.2.7}$$

$$\hat{\mathbf{y}} = [\, y(0), y(1), \ldots , y(N-1) \,]^T \;, \tag{4.2.8}$$

$$\hat{\mathbf{a}} = [\, \hat{A}_1 e^{j\hat{\phi}_1}, \hat{A}_2 e^{j\hat{\phi}_2}, \ldots , \hat{A}_{\hat{M}'} e^{j\hat{\phi}_{\hat{M}'}} \,]^T \;, \tag{4.2.9}$$

where $\hat{s}_i = 2\pi(-\hat{\alpha}_i + j\hat{f}_i)$. The solution vector $\hat{\mathbf{a}}$ can be solved by minimizing any norm (1-norm, 2-norm, ∞-norm, etc.) of the residual vector

$$\mathbf{r} = \hat{\mathbf{y}} - \hat{\boldsymbol{Z}}\hat{\mathbf{a}} \;. \tag{4.2.10}$$

In chapter 5, the $\hat{\mathbf{a}}$'s are estimated by minimizing the 2-norm of the residual vector defined by Equation (4.2.10).

4.2.1 Noiseless case

Now, let us see how the LPSVD method behaves in the noiseless situation. We hope that the method will give us the correct value of the parameters in this situation.

Before we say anything, one might notice that Equations (4.2.1) to (4.2.5) were all written in backward linear prediction instead of the usual forward linear prediction as in Equation (3.1.3). We will justify the reason for using backward linear prediction in this section. Suppose we write the forward linear prediction data matrix $\boldsymbol{A}_F$ and data vector $\mathbf{h}_F$, respectively, as follows :

$$\boldsymbol{A}_F = \begin{bmatrix} s(L-1) & s(L-2) & \cdots & s(0) \\ s(L) & s(L-1) & \cdots & s(1) \\ \vdots & \vdots & & \vdots \\ s(N-2) & s(N-3) & \cdots & s(N-1-L) \end{bmatrix} \qquad (4.2.1.1)$$

$$\mathbf{h}_F = [\, s(L), s(L+1), \ldots\ldots, s(N-1) \,]^T . \qquad (4.2.1.2)$$

Then, the second step of the LPSVD algorithm becomes forming the coefficient vector $\mathbf{g}_F$ of a polynomial of order L by

$$\mathbf{g}_F = -[\boldsymbol{A}_F]_M^+ \mathbf{h}_F . \qquad (4.2.1.3)$$

Now, we will examine the properties of the M signal zeros for both forward and backward linear prediction. Then, we will show the behavior of the L-M extraneous zeros, from that we will see how backward linear prediction is suitable for estimating the parameters for exponentially decaying sinusoidal signals. From the above examination, we can also understand the choice of the AR model order L (i.e., $M \le L \le N-M$) that was mentioned in chapter 3. This section is based mainly on the paper of Kumaresan (1983) and all the theorems will be proved in Appendices 9 to 11.

Theorem 4.2.1.1 Let $s(n) = \sum_{k=1}^{M} a_k e^{s_k n}$, $n = 0, 1, \ldots, N-1$, where a_k and s_k are unknown complex numbers, $s_i \neq s_k$ for $i \neq k$, and

$$\boldsymbol{A}'_F = \begin{bmatrix} s(L) & s(L-1) & \ldots\ldots & s(0) \\ s(L+1) & s(L) & \ldots\ldots & s(1) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ s(N-1) & s(N-2) & \ldots & s(N-L-1) \end{bmatrix} = [\, \boldsymbol{h}_F , \boldsymbol{A}_F \,] . \qquad (4.2.1.4)$$

If the coefficient vector $\mathbf{g}'_F = [\, g_{F,0}, g_{F,1}, g_{F,2}, \ldots, g_{F,L}]^T$ satisfies the homogeneous equation $\mathbf{A}'_F \mathbf{g}'_F = \mathbf{0}$, and if L satisfies the inequality $M \le L \le N-M$, then $G_F(z) = \sum_{k=0}^{L} g_{F,k} z^{-k}$ has M of its L zeros at e^{s_k}, for k = 1, 2, ..., M.

Theorem 4.2.1.2 : Let $s(n) = \sum_{k=1}^{M} a_k e^{s_k n}$, n = 0, 1, ... , N-1, where a_k is the complex amplitude with unknown magnitude A_k and phase ϕ_k, and $s_k/(2\pi)$ is the complex frequency with unknown normalized damping rate α_k and normalized frequency f_k. $s_i \ne s_k$ for $i \ne k$. M is the number of complex sinusoids. If a coefficient vector

$$\mathbf{g}'_B = [\, g_{B,0}, g_{B,1}, g_{B,2}, \ldots, g_{B,L}]^T$$

satisfies the homogeneous equation $\mathbf{A}'_B \mathbf{g}'_B = \mathbf{0}$, and if L satisfies the inequality $M \le L \le N-M$, the polynomial $G_B(z) = \sum_{k=0}^{L} g_{B,k} z^{-k}$ has zeros at $e^{-s_k^*}$, for k = 1, 2, ..., M, where

$$\mathbf{A}'_B = \begin{bmatrix} s^*(0) & s^*(1) & \ldots & s^*(L) \\ s^*(1) & s^*(2) & \ldots & s^*(L+1) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ s^*(N-L-1) & s^*(N-L) & \ldots & s^*(N-1) \end{bmatrix} = [\mathbf{h}_B, \mathbf{A}_B\,] . \tag{4.2.1.5}$$

Basically, both Theorems 4.2.1.1 and 4.2.1.2 say that if $M \le L \le N-M$, we can accurately obtain all the signal zeros back. For exponentially damped sinusoidal signals Theorem 4.2.1.1 states that the signal zeros are within the unit circle if forward linear prediction is used, while Theorem 4.2.1.2 states that the signal zeros are outside the unit circle if backward linear prediction is used. Therefore, if one chooses to use backward linear prediction estimation, the estimated signal zeros needed to be reflected inside the unit circle in order to have the correct sign of the component's damping rate.

What about the L-M extraneous zeros? We have the following condition which can constrain the extraneous zeros to be inside the unit circle.

Theorem 4.2.1.3 : Let $G(z) = 1 + \sum_{k=1}^{L} g_k z^{-k}$, $B(z) = \sum_{k=0}^{M} b_k z^{-k}$, $C(z) = \sum_{k=0}^{L-M} c_k z^{-k}$, and $G(z) = B(z)C(z)$. If $Q = 1 + |g_1|^2 + |g_2|^2 + \ldots + |g_L|^2$ is minimal, and B(z) is given, then the roots of C(z) must be inside the unit circle (see e.g., Lang and McClellan, 1979; Pakula and Kay, 1983; Stoica and Nehorai, 1987).

Notice that in Theorem 4.2.1.3, G(z) can be either forward linear prediction or backward linear prediction. In other words, provided that $M \leq L \leq N-M$ so that the signal zeros are calculated accurately and if $Q = 1 + |g_1|^2 + |g_2|^2 + \ldots + |g_L|^2$ is minimized, then the L-M extraneous roots must not be on nor outside the unit circle no matter which direction of prediction is used.

The coefficient vector $\mathbf{g}'_B = (1, g_{B,1}, g_{B,2}, \ldots, g_{B,L})^T = (1, \mathbf{g}_B^T)$ can be found as follows. Since $\mathbf{g}'_B$ satisfies the homogeneous equations $\boldsymbol{A}'_B \mathbf{g}'_B = \mathbf{0}$ (or $\boldsymbol{A}'_F \mathbf{g}'_F = \mathbf{0}$) and $g_{B,0} = 1$, we can rewrite the homogeneous equations as follows:

$$\boldsymbol{A}_B\, \mathbf{g} = -\mathbf{h}_B\,, \tag{4.2.1.6}$$

because $\boldsymbol{A}'_B$ is partitioned as in Theorem 4.2.1.2 (i.e., $\boldsymbol{A}'_B = [\,\mathbf{h}_B, \boldsymbol{A}_B\,]$) . Hence, from Equation (3.3.6), the unique minimum 2-norm solution of Equation (4.2.1.6) is simply

$$\mathbf{g}_{LS} \;=\; -\boldsymbol{A}_B^{+}\,\mathbf{h}_B \;=\; -\sum_{i=1}^{M}\left(\frac{\mathbf{u}_i^{H}\,\mathbf{h}_B}{\sigma_i}\right)\mathbf{v}_i\;, \tag{4.2.1.7}$$

where $\boldsymbol{A}_B^{+}$ is the pseudo-inverse of $\boldsymbol{A}_B$, σ_i's are the singular values of $\boldsymbol{A}_B$, and the $\mathbf{u}_i$'s and $\mathbf{v}_i$'s are the left and right singular vectors of $\boldsymbol{A}_B$, respectively.

In conclusion, if the LPSVD algorithm is used under the constraints developed in Theorem 4.2.1.1 to Theorem 4.2.1.3, it will estimate the signal parameters accurately in the noiseless situation.

An example to demonstrate the signal zeros and extraneous zeros obtained from LPSVD algorithm under the correct constraints is as follows.

Example 4.2.1 : This example is used to demonstrate Theorems 4.2.1.1-4.2.1.3.

The simulated data are given by the formula

$$y(n) = a_1 e^{s_1 n} + a_2 e^{s_2 n}, \qquad n = 0, 1, 2, \ldots\ldots, 24$$

where

$$a_1 = a_2 = 250,$$

$$s_1 = -50\,\Delta t + j\,2\pi\,(-0.3),$$

$$s_2 = -50\,\Delta t + j\,2\pi\,(-0.275),$$

$$\Delta t = 4 \times 10^{-4} \text{ sec}.$$

Figure 4.2.1.1 shows the zeros of the polynomial G(z) calculated by the LPSVD algorithm for L = 6 and L = 16, respectively. Figures 4.2.1.1(a) and (c) show the result when the forward LPSVD method is used; figures 4.2.1.1(b) and (d) show the result when the backward LPSVD method is used.

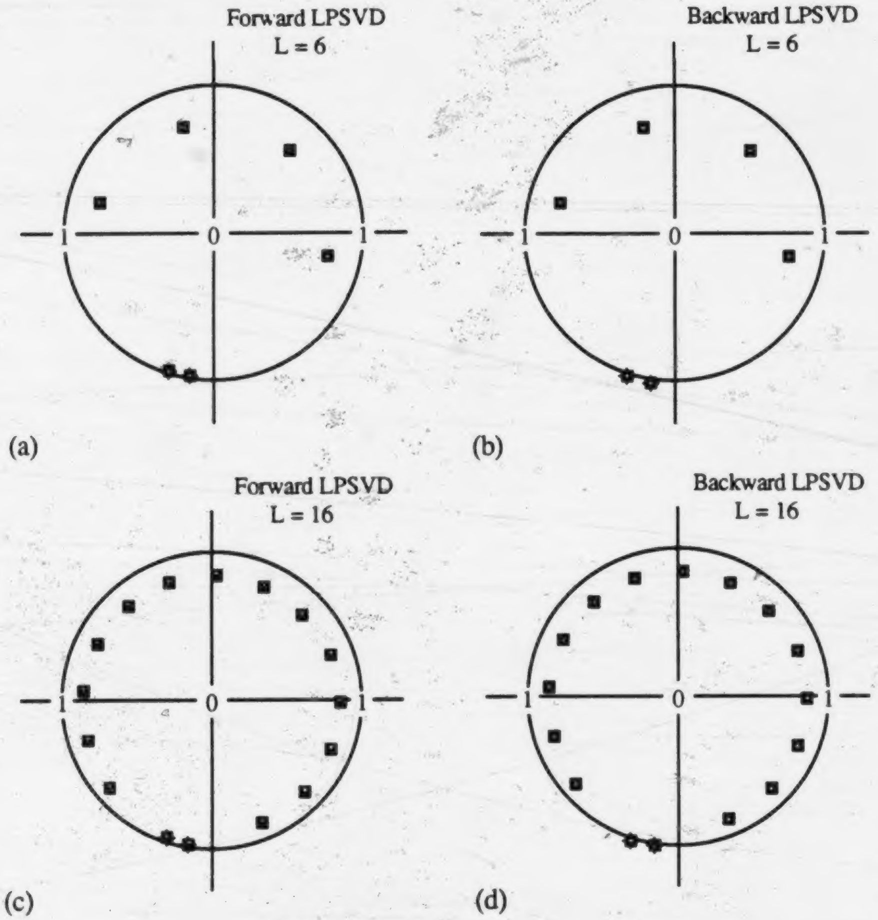


Figure 4.2.1.1 *Demonstration of the zeros of G(z) obtained from the LPSVD algorithm : The zeros of G(z) are superposed with the unit circle. The "+" shows the true locations of the two zeros. If "+"s are inside the unit circle they correspond to* e^{s_1} *and* e^{s_2} *and if they are outside they correspond to* $e^{-s_1^*}$ *and* $e^{-s_2^*}$.

(a) Forward LPSVD : L=6, M=2, N=25 ;
(b) Backward LPSVD : L=6, M=2, N=25 ;
(c) Forward LPSVD : L=16, M=2, N=25 ;
(d) Backward LPSVD : L=16, M=2, N=25 .

4.2.2 Perturbation analysis

In this section, we present the first order perturbation analysis of the LPSVD method used for estimating the normalized damping rates $œ_i$ and normalized frequencies f_i (or $\omega_i = 2\pi f_i/\Delta t$) under relatively small noise. At this moment, perturbation analysis for estimating amplitudes and phases is still under research.

Since for exponentially decaying sinusoids, backward linear prediction provides us a means to distinguish signal zeros from extraneous zeros in the noiseless case, we will consider only backward linear prediction and the subscript "B" is dropped for simplicity. We denote perturbations by preceding the corresponding noisy quantity by Δ. It is assumed that the number of sinusoids M is known and $s_i \neq s_k$ for $i \neq k$. The following theorem is important in our derivation. The proof is given in Appendix 12.

Theorem 4.2.2.1 : Assume

$$\begin{aligned} \hat{\boldsymbol{A}} &= \boldsymbol{A} + \Delta\boldsymbol{A} \\ \hat{\boldsymbol{A}}_M^+ &= \boldsymbol{A}^+ + \Delta\boldsymbol{A}_M^+ , \end{aligned} \tag{4.2.2.1}$$

where $\boldsymbol{A} \in C^{(N-L)xL}$ has rank M. $\Delta\boldsymbol{A}$ is a small perturbation matrix. $\hat{\boldsymbol{A}}_M^+$ is the "truncated rank M" pseudo-inverse of $\hat{\boldsymbol{A}}$. $\boldsymbol{A}^+$ is the pseudo-inverse of $\boldsymbol{A}$ having rank M. $\Delta\boldsymbol{A}_M^+$ is the corresponding perturbation matrix. Then, if $\Delta\boldsymbol{A}$ is small, we have

$$\mathbf{v}_0^H \, \Delta\boldsymbol{A}_M^+ \, \mathbf{u}_0 = - \mathbf{v}_0^H \, \boldsymbol{A}^+ \, \Delta\boldsymbol{A} \, \boldsymbol{A}^+ \, \mathbf{u}_0 , \tag{4.2.2.2}$$

where $\mathbf{v}_0^H$ is any row vector in the row space of $\boldsymbol{A}$; $\mathbf{u}_0$ is any column vector in the column space of $\boldsymbol{A}$.

Using this theorem, we can show that first order perturbations in the estimated zeros, normalized frequencies and normalized damping rates are

$$\Delta z_i \quad = \quad \frac{-1}{|a_i|} \, \frac{\mathbf{p}_i^H \Delta \boldsymbol{A}' \mathbf{g}'}{\sum_{k=1}^{L} k \, g_k \, z_i^{-k-1}} \quad , \qquad (4.2.2.3)$$

$$\Delta f_i \quad = \quad \frac{-1}{2\pi |a_i|} \, \mathrm{Im}\left[\frac{\mathbf{p}_i^H \Delta \boldsymbol{A}' \mathbf{g}'}{\sum_{k=1}^{L} k \, g_k \, z_i^{-k}} \right] \quad , \qquad (4.2.2.4)$$

$$\Delta \alpha_i \quad = \quad \frac{-1}{2\pi |a_i|} \, \mathrm{Re}\left[\frac{\mathbf{p}_i^H \Delta \boldsymbol{A}' \mathbf{g}'}{\sum_{k=1}^{L} k \, g_k \, z_i^{-k}} \right] \quad , \qquad (4.2.2.5)$$

where $\mathbf{p}_i^H$ is the i-th row of the pseudo-inverse of $\boldsymbol{Z}_L$ defined by Equation (4.2.2.11), Im[ ] is the imaginary part of the data in the bracket, Re[ ] is the real part of the data in the bracket, $\mathbf{g}' = [\ 1, \mathbf{g}^T\ ]$, and $\Delta \boldsymbol{A}'$ is matrix filled with noise components, i.e.,

$$\begin{aligned} \Delta \boldsymbol{A}' \quad &= \quad [\ \Delta \mathbf{h}, \Delta \boldsymbol{A}\ ] \\ &= \quad \begin{bmatrix} w^*(0) & w^*(1) & \ldots & w^*(L) \\ w^*(1) & w^*(2) & \ldots & w^*(L+1) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ w^*(N-L-1) & w^*(N-L) & \ldots & w^*(N-1) \end{bmatrix} \quad , \qquad (4.2.2.6) \end{aligned}$$

where the w(k)'s are the noise sequences added in the noiseless FID signal as in Equation (1.1.1).

To show Equation (4.2.2.3), first obtain the first order perturbation of the i-th zero in Equation (4.2.5) (see Appendix 13) which is

$$\Delta z_i = \frac{\sum_{k=1}^{L} \Delta g_k z_i^{-k}}{\sum_{k=1}^{L} k\, g_k z_i^{-k-1}} , \qquad (4.2.2.7)$$

$$= \frac{\mathbf{z}_i^H \Delta\mathbf{g}}{\sum_{k=1}^{L} k\, g_k z_i^{-k-1}} , \qquad (4.2.2.8)$$

where $\mathbf{z}_i^H = [\, z_i^{-1}, z_i^{-2}, \ldots, z_i^{-L} \,]$,

$\Delta\mathbf{g} = [\, \Delta g_1, \Delta g_2, \ldots, \Delta g_L \,]^T$.

Then, obtain the first order perturbation of **g** in Equation (4.2.3) (see Appendix 13) which is

$$\Delta\mathbf{g} = \hat{\mathbf{g}} - \mathbf{g} = -\Delta\boldsymbol{A}_M^{+}\,\mathbf{h} - \boldsymbol{A}^{+}\,\Delta\mathbf{h} . \qquad (4.2.2.9)$$

Now, let us decompose $\boldsymbol{A}$ into

$$\boldsymbol{A} = \boldsymbol{Z}_L\, \boldsymbol{\Lambda}\, \boldsymbol{Z}_R , \qquad (4.2.2.10)$$

where

$$\boldsymbol{Z}_L = \begin{bmatrix} e^{-j\phi_1} & e^{-j\phi_2} & \ldots & e^{-j\phi_M} \\ e^{s_1^*-j\phi_1} & e^{s_2^*-j\phi_2} & \ldots & e^{s_M^*-j\phi_M} \\ e^{2s_1^*-j\phi_1} & e^{2s_2^*-j\phi_2} & \ldots & e^{2s_M^*-j\phi_M} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ e^{(N-L-1)s_1^*-j\phi_1} & e^{(N-L-1)s_2^*-j\phi_2} & \ldots & e^{(N-L-1)s_M^*-j\phi_M} \end{bmatrix} ,$$

(4.2.2.11)

$$\boldsymbol{\Lambda} = \begin{bmatrix} |a_1| & & & & \\ & |a_2| & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & |a_M| \end{bmatrix}, \qquad (4.2.2.12)$$

$$\boldsymbol{Z_R} = \begin{bmatrix} e^{s_1^*} & e^{2s_1^*} & \dots & e^{Ls_1^*} \\ e^{s_2^*} & e^{2s_2^*} & \dots & e^{Ls_2^*} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ e^{s_M^*} & e^{2s_M^*} & \dots & e^{Ls_M^*} \end{bmatrix}. \qquad (4.2.2.13)$$

Since $\mathbf{z}_i^H$ is the i-th row of $\boldsymbol{Z_R}$ defined by Equation (4.2.2.13), it is in the row space of $\boldsymbol{A}$, and $\mathbf{h}$ is a linear combination of columns of $\boldsymbol{Z_L}$ so that it is in the column space of $\boldsymbol{A}$. Therefore,

$$\mathbf{z}_i^H \Delta\mathbf{g} = \mathbf{z}_i^H [- \Delta\boldsymbol{A}_M^+ \mathbf{h} - \boldsymbol{A}^+ \Delta\mathbf{h}] \qquad (4.2.2.14)$$

$$= \mathbf{z}_i^H [\boldsymbol{A}^+ \Delta\boldsymbol{A} \boldsymbol{A}^+ \mathbf{h} - \boldsymbol{A}^+ \Delta\mathbf{h}] \qquad (4.2.2.15)$$

$$= - \mathbf{z}_i^H \boldsymbol{A}^+ \Delta\boldsymbol{A}' \mathbf{g}' . \qquad (4.2.2.16)$$

The right-hand side of Equation (4.2.2.14) is obtained from Equation (4.2.2.9), Equation (4.2.2.15) is obtained from Equation (4.2.2.2), and $\Delta\boldsymbol{A}'$ is defined by Equation (4.2.2.6).

From Equation (4.2.2.10), we have

$$\boldsymbol{A}^+ = \boldsymbol{Z_R}^+ \boldsymbol{\Lambda}^{-1} \boldsymbol{Z_L}^+ \qquad (4.2.2.17)$$

$$= \boldsymbol{Z_R}^H (\boldsymbol{Z_R} \boldsymbol{Z_R}^H)^{-1} \boldsymbol{\Lambda}^{-1} (\boldsymbol{Z_L}^H \boldsymbol{Z_L})^{-1} \boldsymbol{Z_L}^H . \qquad (4.2.2.18)$$

Equation (4.2.2.18) is derived from Theorem 3.2.2 in which the properties of full row rank of $\boldsymbol{Z_R}$ and full column rank of $\boldsymbol{Z_L}$ are used. Since the row rank of $\boldsymbol{Z_R}$ is equal to the rank of $\boldsymbol{Z_R}$,

$$\boldsymbol{Z_R}\,\boldsymbol{Z_R}^{+} \quad = \quad \boldsymbol{I} \ ,$$

where $\boldsymbol{I}$ is an identity matrix. In other words,

$$\mathbf{z}_i^{H}\,\boldsymbol{Z_R}^{+} \quad = \quad [\ \underbrace{0,\ldots,0}_{i\ -1}\,, 1, 0, \ldots\ldots, 0\ ]\ . \qquad (4.2.2.19)$$

Let $\mathbf{p}_i^{H}$ be the i-th row of $\boldsymbol{Z_L}^{+}$. Then, from Equation (4.2.2.17),

$$\begin{aligned} \mathbf{z}_i^{H}\,\boldsymbol{A}^{+} \quad &= \quad \mathbf{z}_i^{H}\,\boldsymbol{Z_R}^{+}\,\boldsymbol{\Lambda}^{-1}\,\boldsymbol{Z_L}^{+} \\ &= \quad \frac{1}{|a_i|}\,\mathbf{p}_i^{H}\ . \qquad (4.2.2.20) \end{aligned}$$

Combining Equations (4.2.2.20), (4.2.2.16) and (4.2.2.8), we can prove Equation (4.2.2.3).

To show Equation (4.2.2.4), notice that

$$\begin{aligned} 2\pi\,\hat{f}_i \quad &= \quad \mathrm{Im}\,[\ \ln(\hat{z}_i)\ ] \\ &= \quad \mathrm{Im}\,[\ \ln(\ z_i + \Delta z_i\ )\ ] \\ &\approx \quad \mathrm{Im}\left[\ln(z_i) + 2\left[\frac{\Delta z_i}{2z_i + \Delta z_i}\right] + \ldots.\right]\ . \end{aligned}$$

Therefore, the first order perturbation of f_i may be written as

$$\Delta f_i \quad = \quad \frac{1}{2\pi}\,\mathrm{Im}\left[\frac{\Delta z_i}{z_i}\right]\ . \qquad (4.2.2.21)$$

Equation (4.2.2.5) can be shown similarly.

We assume that the additive noise w(k) is zero mean white noise, with the real and imaginary part uncorrelated, and has variance σ_W^2 for each part. In other words,

$$E\{w(k)\} = 0 , \tag{4.2.2.22}$$

$$E\{w(k)w(l)\} = 0 , \tag{4.2.2.23}$$

$$E\{w(k)w^*(l)\} = 2\,\sigma_W^2\,\delta_{k,l} , \tag{4.2.2.24}$$

where $\delta_{k,l}$ is the Kronecker delta function. Then, one can verify from Equation (4.2.2.3) that the mean and variance of Δz_i are

$$E(\Delta z_i) = 0 , \tag{4.2.2.25}$$

$$\mathrm{Var}(\Delta z_i) = \frac{2\sigma_W^2}{|a_i|^2}\,\frac{p_i^H \boldsymbol{R_g}\, p_i}{\left|\sum_{k=1}^{L} k\, g_k\, z_i^{-k-1}\right|^2} , \tag{4.2.2.26}$$

where

$$(\boldsymbol{R_g})_{i,j} = (\boldsymbol{R_g})_{i-j} = (\boldsymbol{R_g})^*_{j,i} = \begin{cases} \sum_{k=i-j}^{L} g_k\, g^*_{k-(i-j)} \;; & 0 \le i\text{-}j \le L \\ 0 \;; & \text{otherwise.} \end{cases} \tag{4.2.2.27}$$

The proof is given in Appendix 13 .

From Equation (4.2.2.21), we can find the variance of Δf_i and $\Delta \alpha_i$ as follows:

$$\begin{aligned} \mathrm{Var}(\Delta f_i) &= \frac{1}{4\pi^2}\, E\left\{ \left[\mathrm{Im}\left(\frac{\Delta z_i}{z_i} \right) \right]^2 \right\} \\ &= \frac{1}{2}\,\frac{1}{4\pi^2}\, E\left\{ \left(\frac{\Delta z_i}{z_i} \right)^2 \right\} \end{aligned}$$

$$= \frac{1}{2}\,\frac{1}{4\pi^2}\,\text{Var}\left(\frac{\Delta z_i}{z_i}\right). \qquad (4.2.2.28)$$

The second equality is obtained by assuming (or approximating)

$$\text{Var}\left\{\text{Re}\left(\frac{\Delta z_i}{z_i}\right)\right\} = \text{Var}\left\{\text{Im}\left(\frac{\Delta z_i}{z_i}\right)\right\} = \frac{1}{2}\,\text{Var}\left(\frac{\Delta z_i}{z_i}\right). \qquad (4.2.2.29)$$

Similarly, we can show that

$$\text{Var}(\Delta \text{œ}_i) = \frac{1}{2}\,\frac{1}{4\pi^2}\,\text{Var}\left(\frac{\Delta z_i}{z_i}\right) = \frac{\sigma_W^2}{4\pi^2 |a_i|^2}\,\frac{\mathbf{p}_i^H \boldsymbol{R}_g\,\mathbf{p}_i}{\left|\sum_{k=1}^{L} k\,g_k\,z_i^{-k}\right|^2}\,. \qquad (4.2.2.30)$$

A number of other properties can also be developed from the above observations.

Property 1 : $\text{Var}(\Delta z_i)$, $\text{var}(\Delta f_i)$ and $\text{var}(\Delta \text{œ}_i)$ are invariant of $|a_j|$ for $j \neq i$, but proportional to $\frac{1}{|a_i|^2}$.

Property 2 : $\text{Var}(\Delta z_i)$, $\text{var}(\Delta f_i)$ and $\text{var}(\Delta \text{œ}_i)$ are invariant to shifting all frequencies by a constant value and/or changing all phases by another constant.

Property 1 can be shown easily using Equations (4.2.2.26), (4.2.2.28) and (4.2.2.29). The proof of property 2 is given in Appendix 14.

In chapter 5, we will compare examples of $\text{var}(\Delta f_i)$ and $\text{var}(\Delta \text{œ}_i)$ to their corresponding Cramer-Rao (CR) lower bound.

4.3 A proposed method

Although from Equation (4.2.2.25), first order perturbation analysis of the LPSVD algorithm indicated that estimates of the parameters are unbiased, simulation studies seem to show that the estimators are biased and their variances may depart from the lowest variances given by the CR bounds.

Here, we propose to use estimates from the LPSVD algorithm as initial guesses for any standard nonlinear iterative algorithm which minimizes the 2-norm of the residual vector as given by Equation (1.4.1.4), or equivalently,

$$S \quad = \quad \sum_{n=0}^{N-1} | \, y(n) - \hat{y}(n) \, |^2 \qquad (4.3.1)$$

with respect to A_i, ϕ_i, œ_i, and f_i ($\omega_i = 2\pi f_i/\Delta t$), for i = 1, 2, . . ., $\hat{M}$, where

$$\hat{y}(n) \quad = \quad \sum_{k=1}^{\hat{M}} \hat{A}_k \, e^{j\hat{\phi}_k} \, e^{2\pi(-\hat{\text{œ}}_k + j\hat{f}_k)n} \, . \qquad (4.3.2)$$

Hopefully by doing this, we can improve the estimates.

As shown in section 1.4.1, if the M noiseless exponentially decaying sinusoidal signals are embedded in a sequence of independent, zero mean, complex valued, Gaussian random noise, then the maximum likelihood estimate of the parameters A_i, ϕ_i, œ_i, and f_i ($\omega_i = 2\pi f_i/\Delta t$), for i = 1, 2, . . ., M, is equivalent to solving Equation (4.3.1). Coarse searches that look for the minimum of the objective function S evaluated for a range of values of A_i, ϕ_i, œ_i, and f_i can be used to obtain the initial estimates. But such techniques are unsuitable for search in many dimensions, and convergence to the global minimum is not guaranteed. Therefore, we propose to use estimates from the LPSVD algorithm as initial guesses for a minimization method that minimizes Equation (4.3.1).

The minimization method used in this thesis is basically the Levenberg-Marquardt method (Appendix 21, Press, et al., 1986, p.521-528). The only major difference is that we modified the algorithm so that the updated parameters have the following constraints :

$$\begin{aligned} \hat{A}_i &\geq 0\,; \\ \hat{\text{œ}}_i &\geq 0\,, \\ -0.5 \leq \hat{f}_i &\leq 0.5, \end{aligned}$$

for i = 1, 2, ... , $\hat{M}$.

Chapter 5

Simulation results

5.1 Introduction

In this chapter, we provide a quantitative comparison of the performance of the following three algorithms: the discrete Fourier transform (DFT) method, the LPSVD algorithm and the proposed algorithm discussed in section 4.3. By the DFT method, we simply mean the discrete Fourier transformation of the time domain FID data. To simplify comparisons, we will not include any of the techniques such as zero-filling, convolution difference, etc., that were mentioned in sections 2.8.2 to 2.8.6, with the only exception being that in section 5.9 we will use the spline function for baseline correction. To further simplify comparisons, we will study spectra that have no phase problem, so that no phase correction will be required for the DFT method.

In order to provide a better understanding of the performance of the algorithms, we have measured the bias (mean of the estimated value - correct input value), the variance and the mean-squared error (MSE, which means, variance + the square of the bias) of the parameter estimates. The variance and mean-squared error of the estimates are compared to their Cramer-Rao (CR) bounds, if possible. The CR bound specifies the lowest bound on the variance of unbiased parameter estimates. This is because for an unbiased estimator $\Lambda'(v_0, v_1, \ldots, v_{N-1})$, where the dependence of the estimator on the observations $v_0, v_1, \ldots, v_{N-1}$ has been explicitly indicated,

$$\sigma^2_{\Lambda'} \geq \frac{1}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\frac{d \ln p(\mathbf{v}|\Lambda)}{d\Lambda}\right]^2 p(\mathbf{v}|\Lambda)\, d\mathbf{v}} , \tag{5.1.1}$$

where $\sigma_{\Lambda'}^2$ is the variance of the unbiased estimator Λ',

$\mathbf{v} \equiv (v_0, v_1, \ldots, v_{N-1})$,

$d\mathbf{v} = dv_0\, dv_1 \ldots dv_{N-1}$,

$p(\mathbf{v}|\Lambda)$ is the probability density of obtaining the N data samples given that the true parameter is Λ.

The right-hand side of Equation (5.1.1) is defined as the CR lowest bound. These bounds serve as a goal for our estimation algorithms. Proof of the existence of CR bounds can be found in Fante (1988, p. 398). Derivation of the CR bounds for the parameters of M exponentially decaying signals in white Gaussian noise is given in Appendix 15.

A summary of the parameters to be compared is given in Table 5.1.

	DFT spectrum	the LPSVD method	the proposed method	perturbation theory	CR lower bounds
ratio of areas from the real part of the spectrum	√	√	√		
amplitude	√	√	√		√
phase	√	√	√		√
damping rate	√	√	√	√	√
frequency	√	√	√	√	√

***Table 5.1** Parameters that will be compared by different methods in this chapter*

The ratio of the peak areas from the real part of the 'phase corrected' spectrum is included because some NMR spectroscopists are interested in this parameter when comparison with a concentration standard is not possible.

The perturbation theory values for the normalized frequencies of M exponentially decaying sinusoids embedded in complex white Gaussian noise are given by Equation (4.2.2.28), while their normalized damping rates are given by Fquation (4.2.2.30).

We have not yet derived the CR bounds for the ratio of the areas of the peaks from the real part of the spectrum, so we can only use the CR lower bounds to compare the variance and MSE of other parameters. The CR bounds for the amplitudes, phases, normalized damping rates and normalized frequencies of M exponentially decaying signals in white Gaussian noise are given in Equations (A15.8), (A15.9), (A15.10) and (A15.11), respectively.

The relationship of the amplitude and phase of an exponentially decaying signal in the time domain to the area under the real part of its spectrum is given in Appendix 17.

The relationship of the damping rate of an exponentially decaying signal to the linewidth of the corresponding peak in the frequency domain for the real part of the spectrum with no phase error is given in Appendix 18. The damping rate estimated from the DFT method is obtained from the measured linewidth of the peak.

In section 5.2, we will study the effect of the tentative model order L and the available number of data points N defined in Equation (4.2.1). In section 5.3, we will study the effect of choosing a different cut-off number of singular values $\hat{M}$ defined in Equation (4.2.3). Since L and $\hat{M}$ are not related to the DFT based algorithm, the results in these two sections will not be compared with DFT spectrum analysis.

Starting from section 5.4, where we will study the effect of noise, we will compare the parameters as tabulated in Table 5.1. In section 5.5, the effect of peak frequency separation will be evaluated from well separated to partially overlapping peaks. The effect of changing a peak's damping rate will be studied in section 5.6, where we will see how the three algorithms perform under the problem produced by overlapping Lorentzian peaks. In section 5.7, we will study the effect of changing one peak's amplitude, while the effect of adding a well separated narrow peak to a spectrum will be discussed in section 5.8. Finally, in section 5.9, we will study the effect of variable baseline in the frequency domain.

For the following sections, there is always at least one peak in the spectrum to be analyzed. The one located at normalized frequency[1] -0.3 is the reference peak which will not be changed within an example. We numbered this peak as the first peak. The other peaks in the spectrum are numbered in order of increasing frequencies.

5.2 Number of tentative model order and data points for the LPSVD method and the proposed method

In this section, we examine three examples to show the effect of changing the tentative model order L defined in Equation (4.2.1) and the available number of data points N on the performance of the LPSVD method and the proposed method.

Example 5.2.1 :

We assume that there are two signals present ($M = 2$), sampled at $\Delta t = 4*10^{-4}$ sec (or 2500 Hz), the second at a normalized frequency of -0.275. The number of data points is N = 50. The normalized frequency difference is $f_1 - f_2 = -0.3 - (-0.275) = -0.025$. The normalized damping rates α_1 and α_2 are 0.00318 for both signals (which is equivalent to $0.00318/\Delta t$ = 7.96 Hz or 50 rad/sec). The amplitude and phase are, respectively, 250 and 0 for both signals ($A_1 = A_2 = 250$; $\phi_1 = \phi_2 = 0$). Sequences of white complex Gaussian noise, with independent real and imaginary parts, each having a variance of $\sigma_W^2 = (12.5)^2 = 156.25$ are added to the sum of the two signals.

---

[1] The relationship between normalized frequency and frequency in units of ppm is shown in Appendix 3. For example, if the spectrometer frequency is 100 MHz and the sampling rate is 2500 Hz, then the normalized frequency of -0.3 is equivalent to -7.5 ppm. On the other hand, if the spectrometer frequency is 40.55 MHz, then the normalized frequency -0.3 is equal to -18.496 ppm.

To give a feeling of where the peaks are located in the spectrum, a representative 1024 point DFT spectrum is shown in Figure 5.2.1.1.

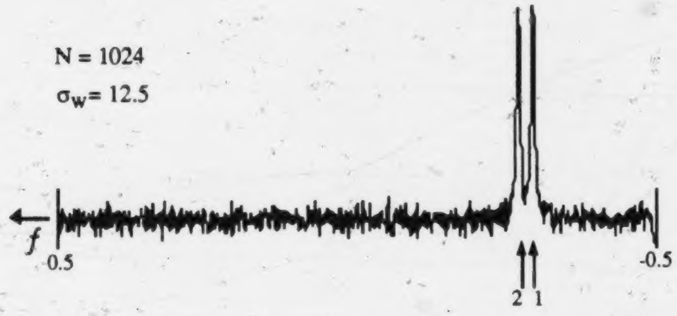


Figure 5.2.1.1 A representative 1024-point DFT spectrum for Examples 5.2.1 to 5.2.3 consisting of two time domain sinusoids of equal amplitude (250), equal damping rate (0.00318) and of normalized frequency separation 0.025. σ_W is the noise standard deviation in time domain defined in Equation (1.1.1).

Figures 5.2.1.2 - 5.2.1.6 show the results of the estimates of the performance measures of the spectral parameters listed in Table 5.1. The straight lines are the CR lower bounds. The "dotted" lines are computed from perturbation analysis. The "squares" and "diamonds" are the results from the LPSVD method and the

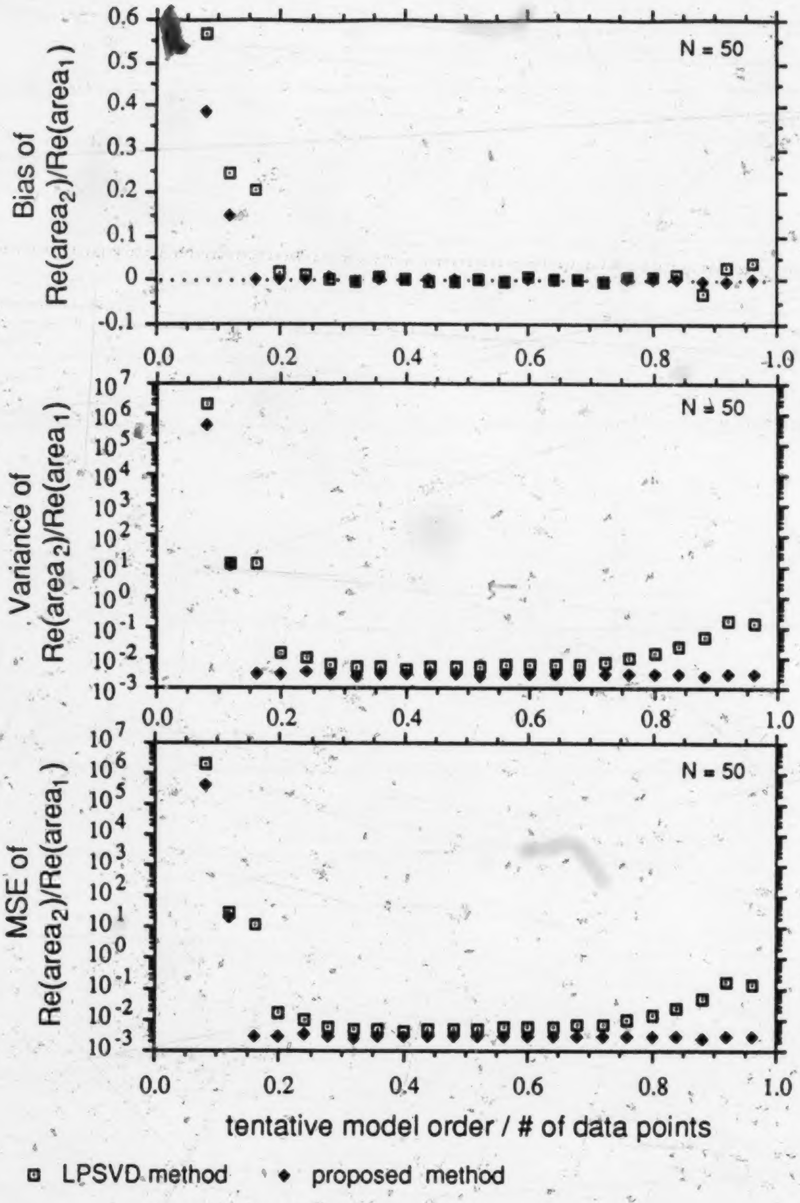


Figure 5.2.1.2 The area ratio of the two peaks from the real spectrum versus the ratio of the tentative model order L to the number of data points N. N=50; $\Delta t = 4 \times 10^{-4}$ sec; $A_1 = A_2 = 250$; $\phi_1 = \phi_2 = 0$; $\text{œ}_1 = \text{œ}_2 = 0.00318$; $f_1 - f_2 = -0.3 - (-0.275) = -0.025$. Each point was obtained from over 400-run simulations. Bias and MSE are defined in section 5.1.

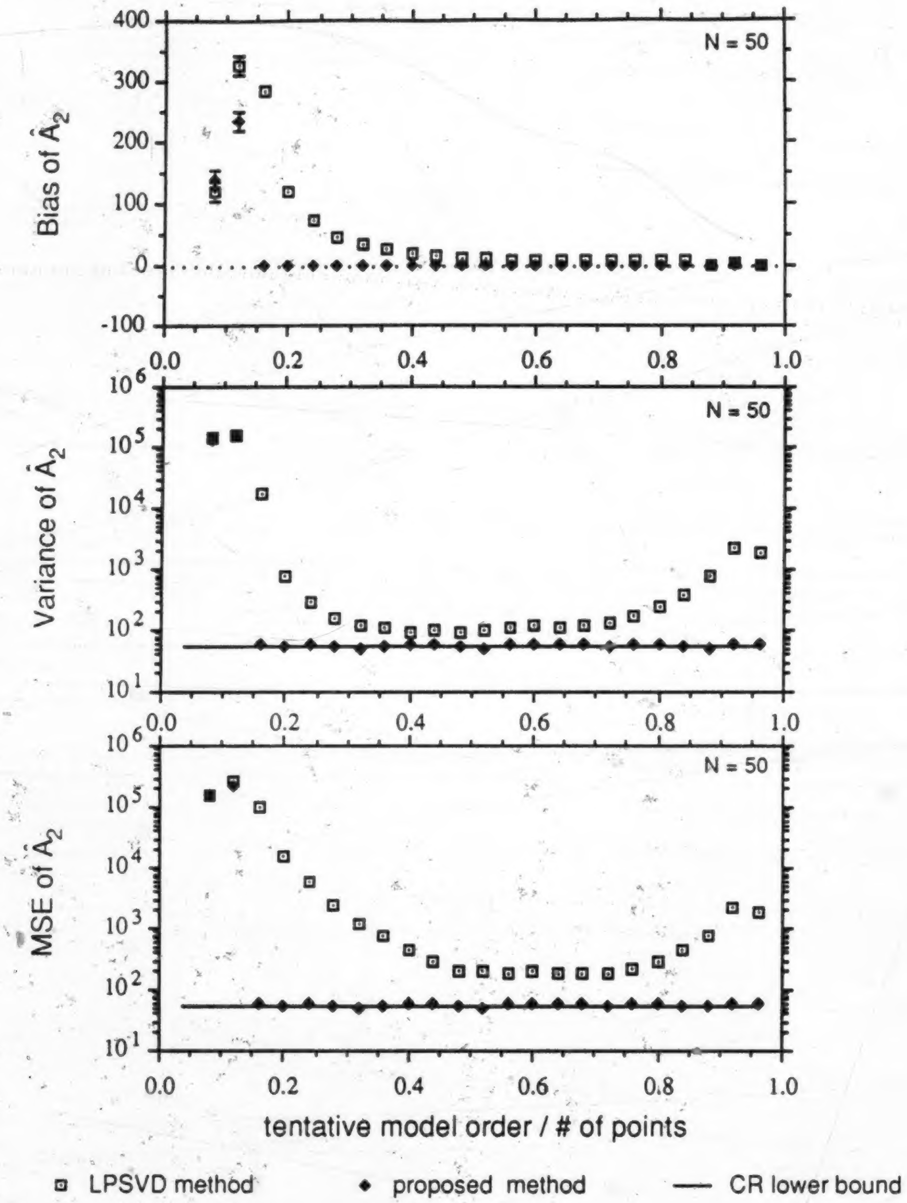


***Figure** 5.2.1.3 Time domain amplitude of the 2-nd peak versus the ratio of the tentative model order to the number of data points. For other information, see Figure 5.2.1.2.*

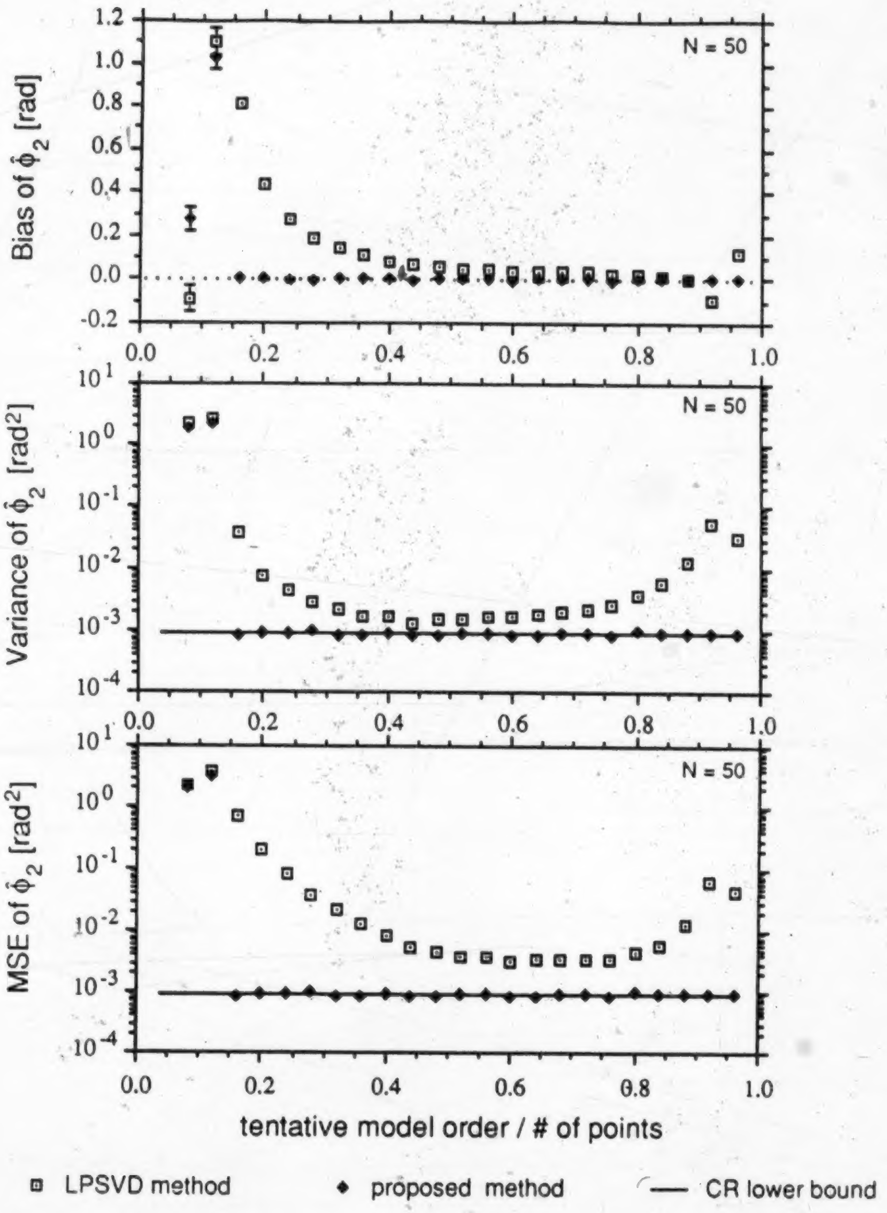


__Figure 5.2.1.4__ Phase of the 2-nd peak versus the ratio of the tentative model order to the number of data points. For other information, see Figure 5.2.1.2.

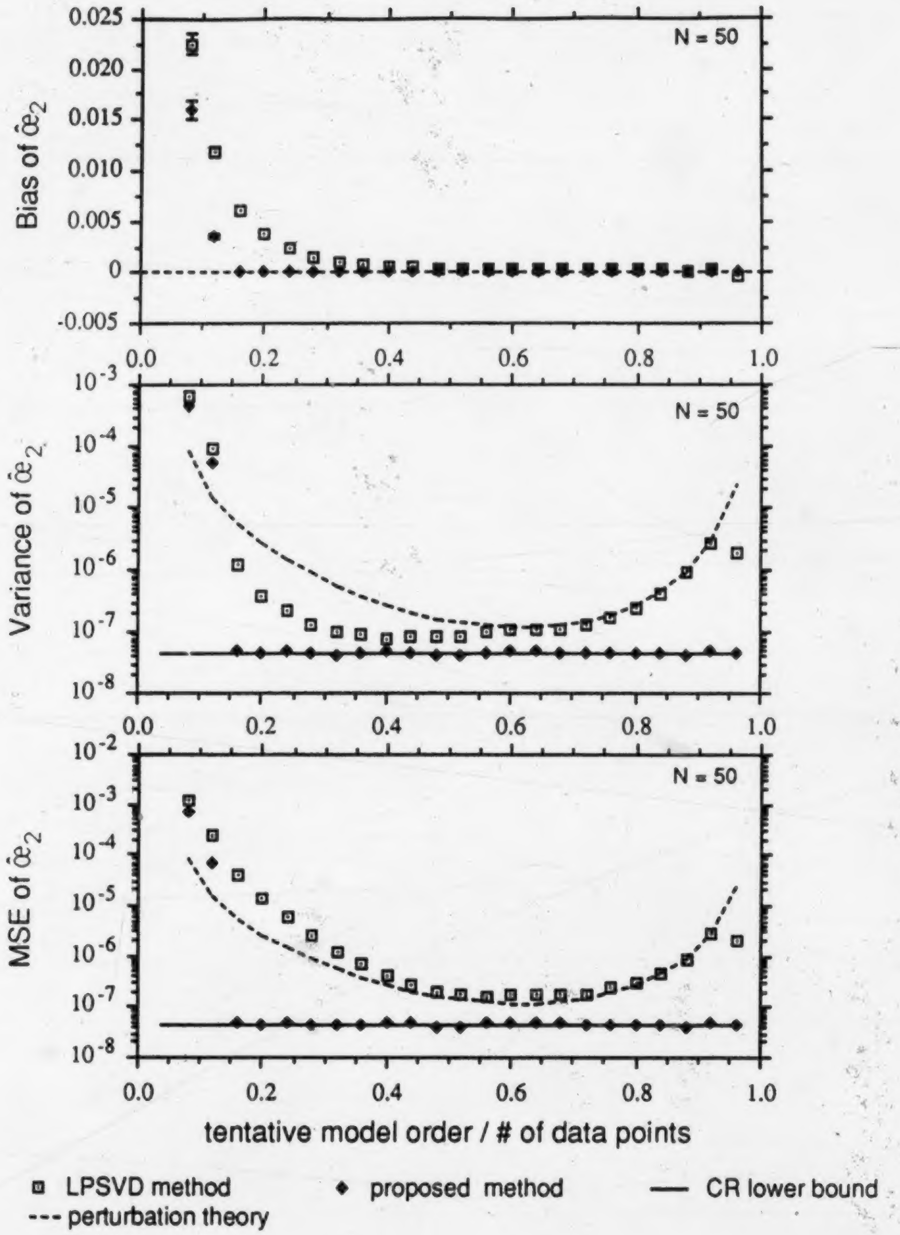


Figure 5.2.1.5 *Normalized damping rate of the 2-nd peak versus the ratio of the tentative model order to the number of data points. For other information, see Figure 5.2.1.2.*

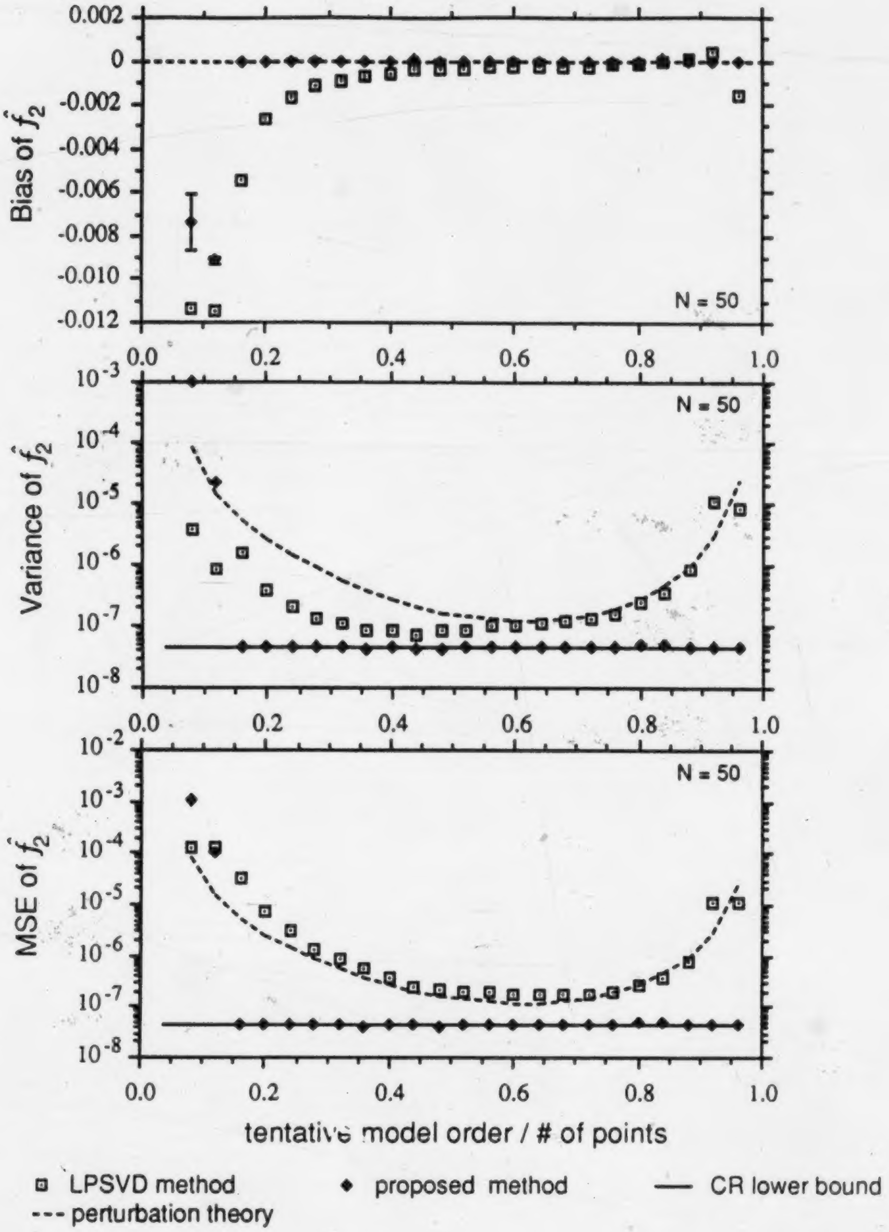


Figure 5.2.1.6 Normalized frequency of the 2-nd peak versus the ratio of the tentative model order to the number of data points. For other information, see Figure 5.2.1.2.

proposed method, respectively. Each estimate was computed from at least 400 independent runs, each with different sequences of noise.

As we can see, the proposed method performs better than the LPSVD method for L > 0.16N = 8 in this example. If we define the optimal choice of L for each parameter to be in the range where the parameter has the smallest MSE, then the optimal ranges of L for each parameter obtained from the LPSVD algorithm are shown in Table 5.2.1.1.

For the proposed method, the optimal range of L is extended from 0.16N to 0.96N.

optimal range of L	MSE (xN)
area ratio	0.24 - 0.72
amplitude A_2	0.44 - 0.80
phase ϕ_2	0.48 - 0.80
normalized damping rate œ_2	0.40 - 0.80
normalized frequency f_2	0.44 - 0.80

***Table 5.2.1.1** Optimal range of the tentative model L for the LPSVD algorithm with N = 50.*

Example 5.2.2 :

This example is designed to test the performance of the LPSVD method and the proposed method for a different number of data points while at the same time allowing L to vary. The same assumptions as in Example 5.2.1 hold, except that N = 100.

Figures 5.2.2.1 to 5.2.2.5 show the results of the estimates of the performance measures of the spectral parameters. In order to facilitate comparison with the last example, the results from the previous example are displayed with the current results. The lighter "squares" and "diamonds" are the results from the previous example for the LPSVD method and the proposed method, respectively.

Again, we see that the proposed method performed better than the LPSVD method. If we define the optimal choice of L for each parameter to be in the range where the parameter has the smallest MSE, then the optimal ranges of L for each parameter obtained from the LPSVD algorithm are shown in Table 5.2.2.1.

optimal range of L	MSE (xN)
area ratio	0.12 - 0.92
amplitude A_2	0.36 - 0.80
phase ϕ_2	0.36 - 0.68
normalized damping rate œ_2	0.40 - 0.80
normalized frequency f_2	0.36 - 0.76

***Table 5.2.2.1** Optimal range of the tentative model L for the LPSVD algorithm with N = 100.*

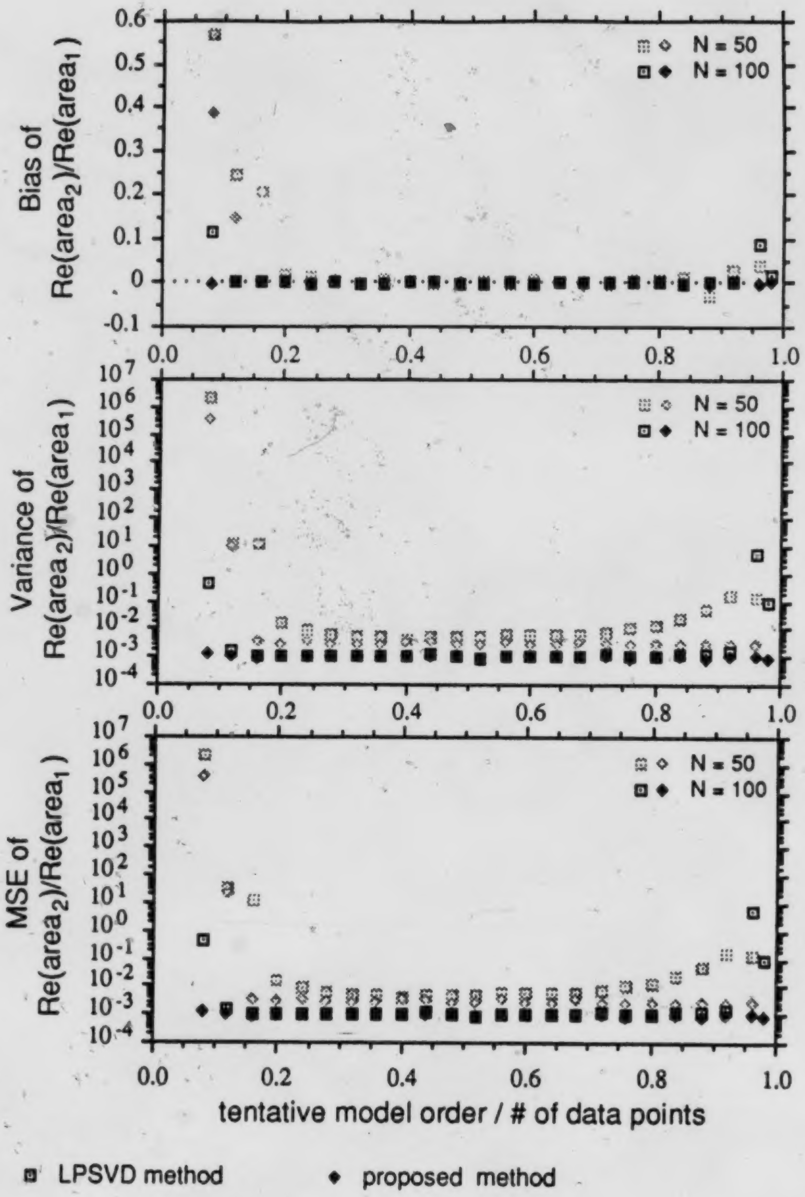


Figure 5.2.2.1 The area ratio of the two peaks from the real spectrum versus the ratio of the tentative model order L to the number of data points N. $\Delta t = 4 \times 10^{-4}$ sec; $A_1 = A_2 = 250$; $\phi_1 = \phi_2 = 0$; $\alpha_1 = \alpha_2 = 0.00318$; $f_1 - f_2 = -0.3 - (-0.275) = -0.025$. Each point was obtained from over 400-run simulations. The lighter squares and diamonds are the results of Example 5.2.1. Bias and MSE are defined in section 5.1.

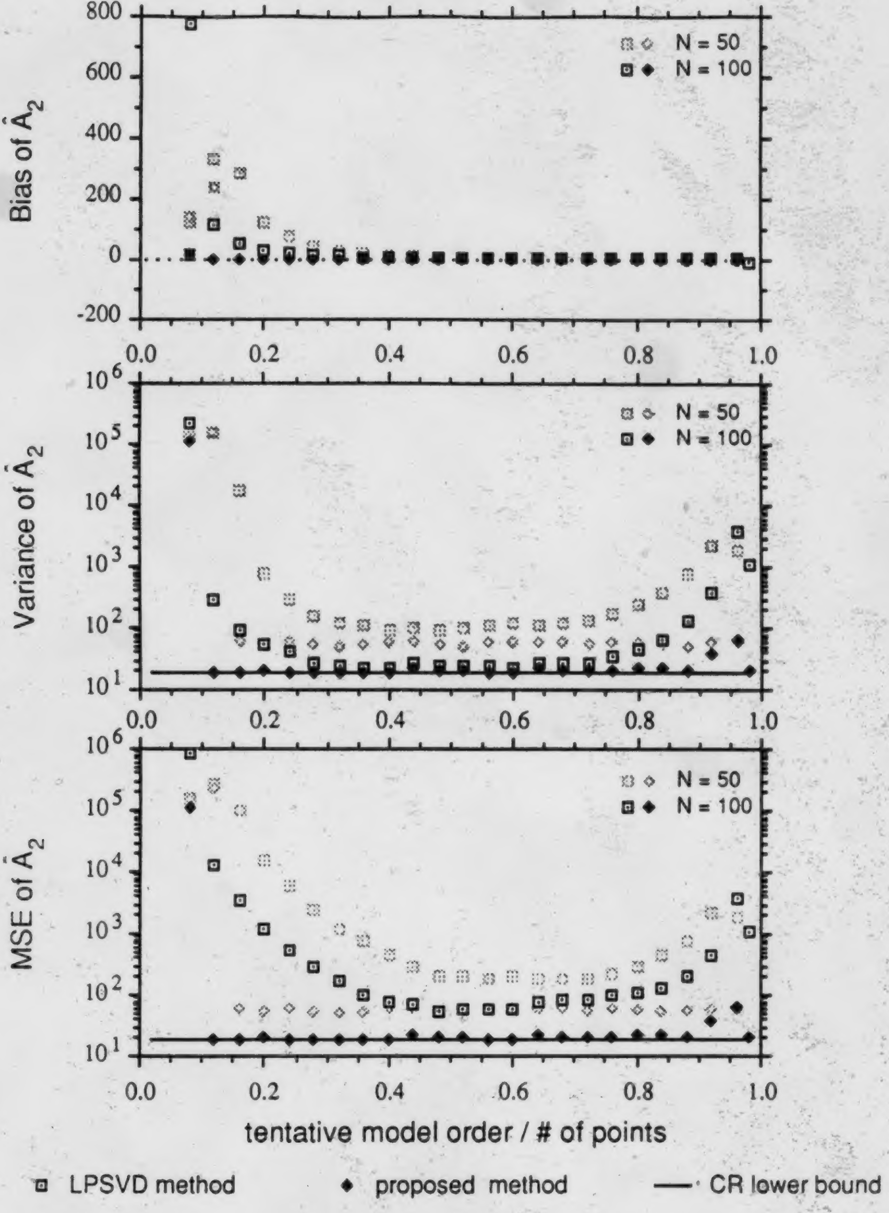


Figure 5.2.2.2 Time domain amplitude of the 2-nd peak versus the ratio of the tentative model order to the number of data points. For other information, see Figure 5.2.2.1.

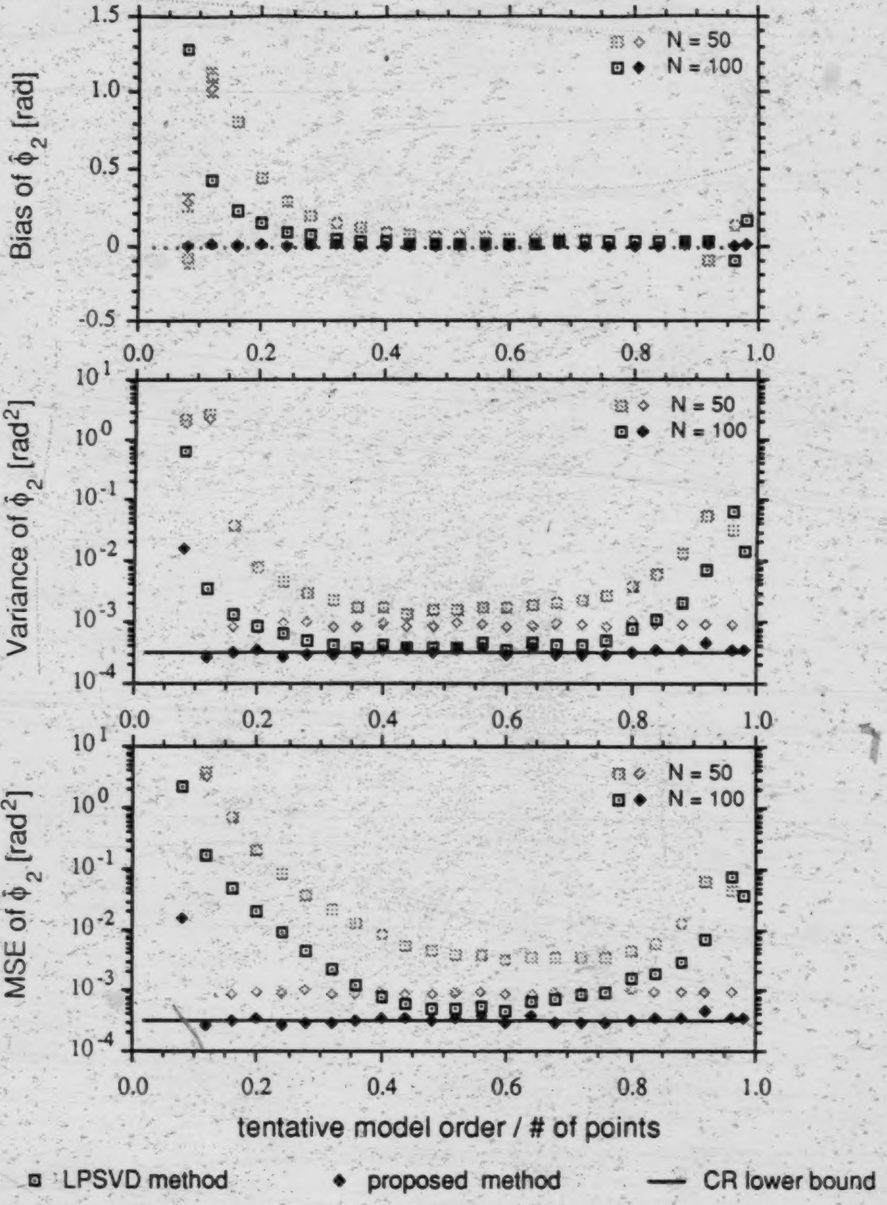


Figure 5.2.2.3 Phase of the 2-nd peak versus the ratio of the tentative model order to the number of data points. For other information, see Figure 5.2.2.1.

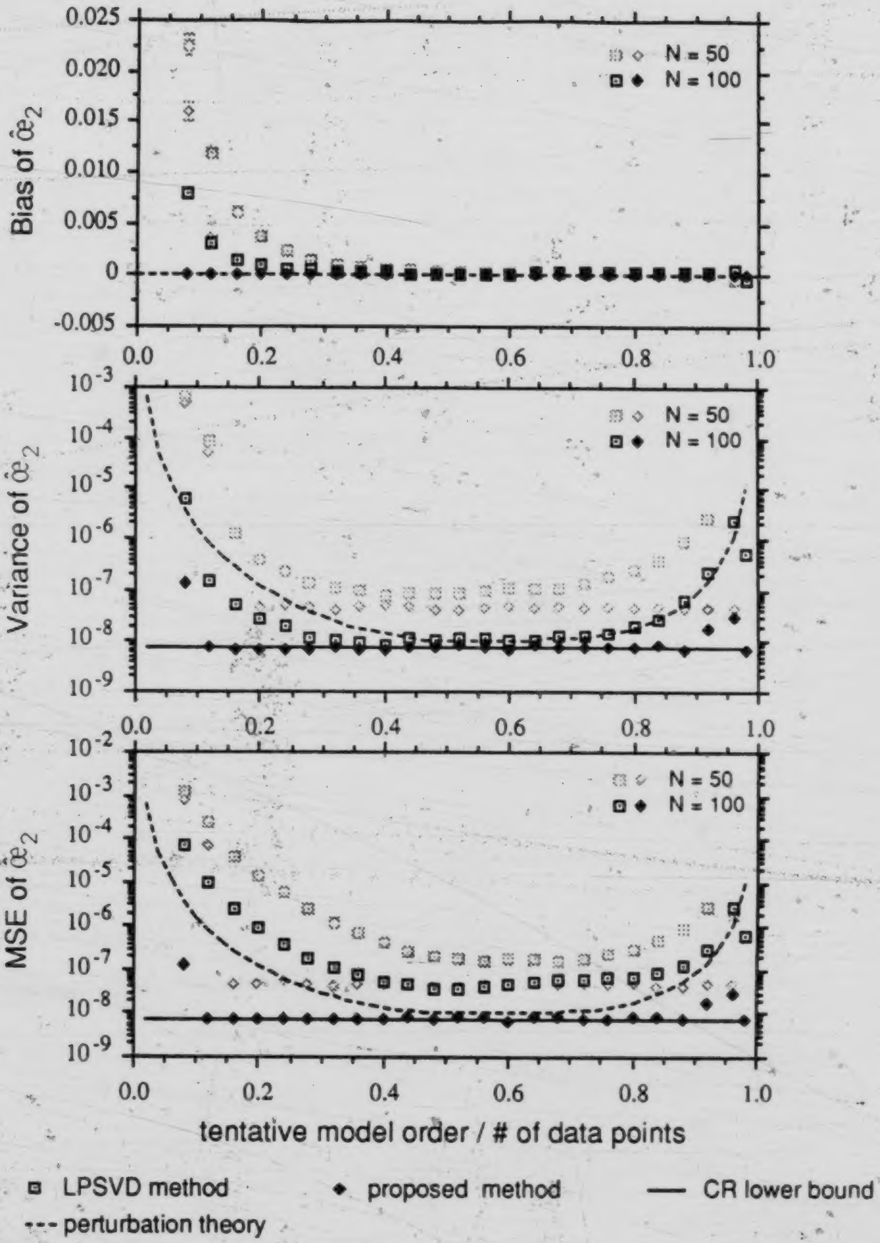


Figure 5.2.2.4 Normalized damping rate of the 2-nd peak versus the ratio of the tentative model order to the number of data points. For other information, see Figure 5.2.2.1.

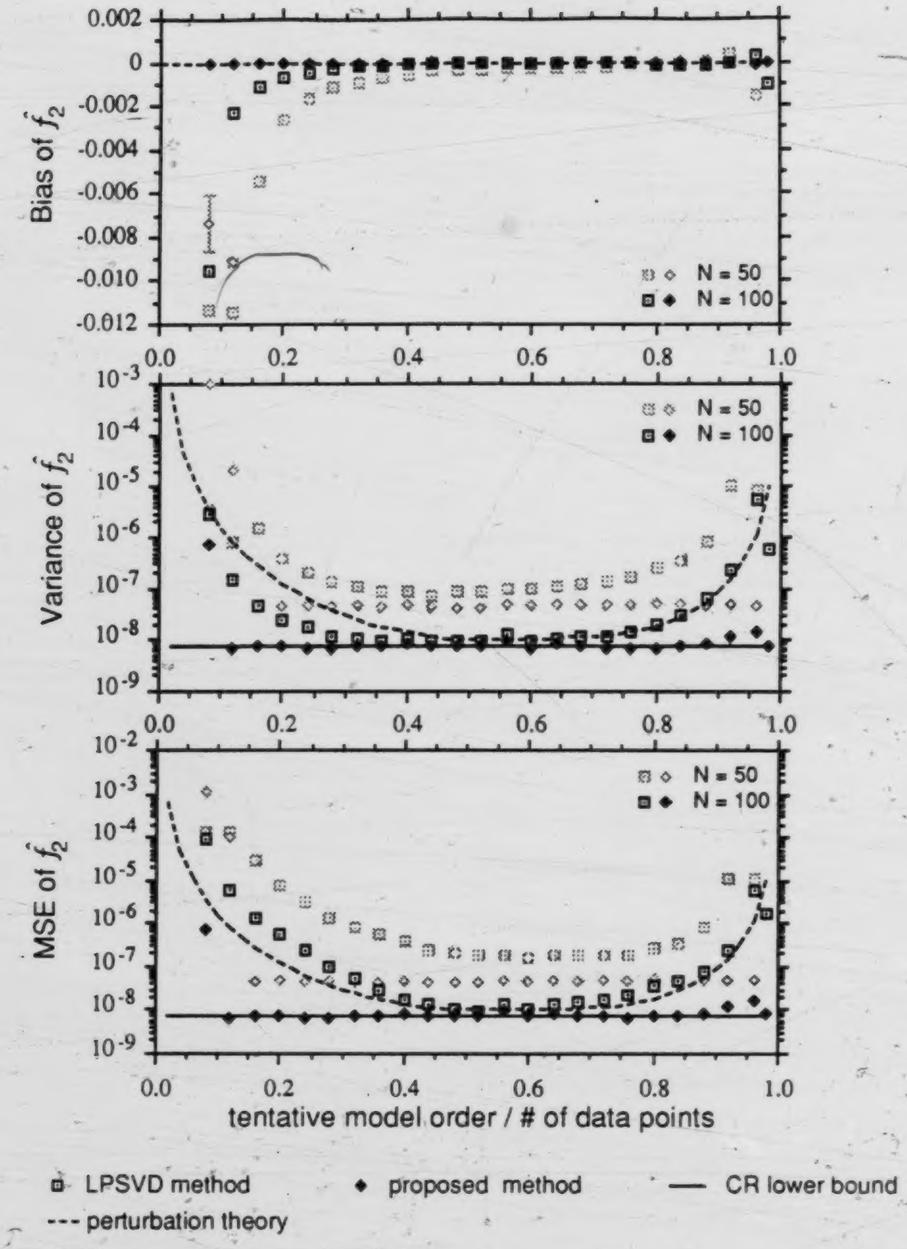


Figure 5.2.2.5 *Normalized frequency of the 2-nd peak versus the ratio of the tentative model order to the number of data points. For other information, see Figure 5.2.2.1.*

For the proposed method, the optimal range of L is extended from 0.12N to 0.98N.

Comparing the results for N = 50 and those for N = 100, we see clearly that increasing the number of data points improves the performance of both the LPSVD method and the proposed method.

From Figures 5.2.2.1 to 5.2.2.5, we notice that not only the ratio of the tentative model order to the number of data points available affects the performance of the LPSVD and the proposed algorithm, but also the available number of data points itself. The following example is designed to observe the effect of the available number of data points N on the performance of the two methods when the ratio of the tentative model order L to the number of data points N is kept constant.

Example 5.2.3 :

The same assumptions as in Example 5.2.1 hold, except that N is varied from 25 to 225, while the tentative model order L is kept equal to 0.52N in all cases, which is a good choice from the last two examples. At least six hundred independent runs were used for each value of N.

Figure 5.2.3.1 shows the statistical result of the first three singular values from the LPSVD method for different number of points N. Figures 5.2.3.2 to 5.2.3.6 show the results of the performance measures of the spectral parameters. Again, we see that the proposed method performs like the best unbiased estimator whose performance is predicted by the CR lower bounds. As predicted by perturbation theory in section 4.2.2 and the CR bounds in Appendix 15, the

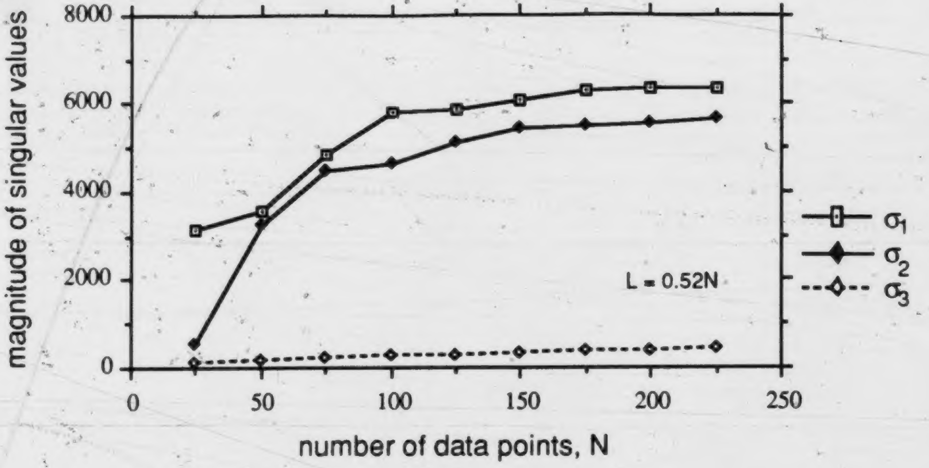


Figure 5.2.3.1 Statistical result of the magnitude of the first three singular values of the LPSVD method for different numbers of data points N. The tentative model order L is varied so that L = 0.52N. Time domain noise standard deviation σ_W is 12.5.

performance of the LPSVD method and the proposed method improve as the number of data points increases, and finally "levels off" at N ≥ 200 data points. The threshold value of N, below which both methods start to deteriorate significantly, is about 50. This threshold value is especially pertinent for the estimate of the ratio of areas of the real part of the spectrum.

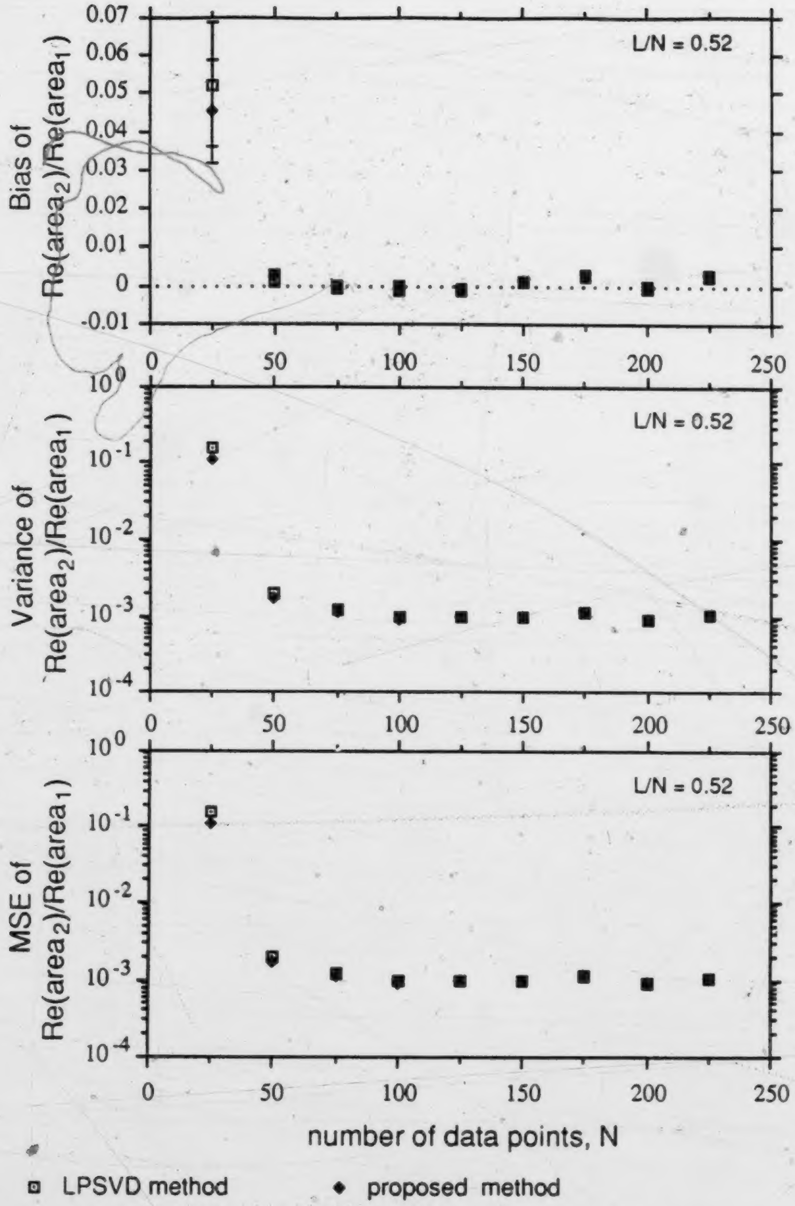


Figure 5.2.3.2 The area ratio of the two peaks from the real spectrum versus number of data points N. L is the tentative model order defined in Equation (4.2.1); N = 25 to 225; $\Delta t = 4 \times 10^{-4}$ sec; $A_1 = A_2 = 250$; $\phi_1 = \phi_2 = 0$; $\alpha_1 = \alpha_2 = 0.00318$; $f_1 - f_2 = -0.3 - (-0.275) = -0.025$. The squares and diamonds, for the LPSVD method and the proposed method, respectively, were obtained from over 600-run simulations. Bias and MSE are defined in section 5.1.

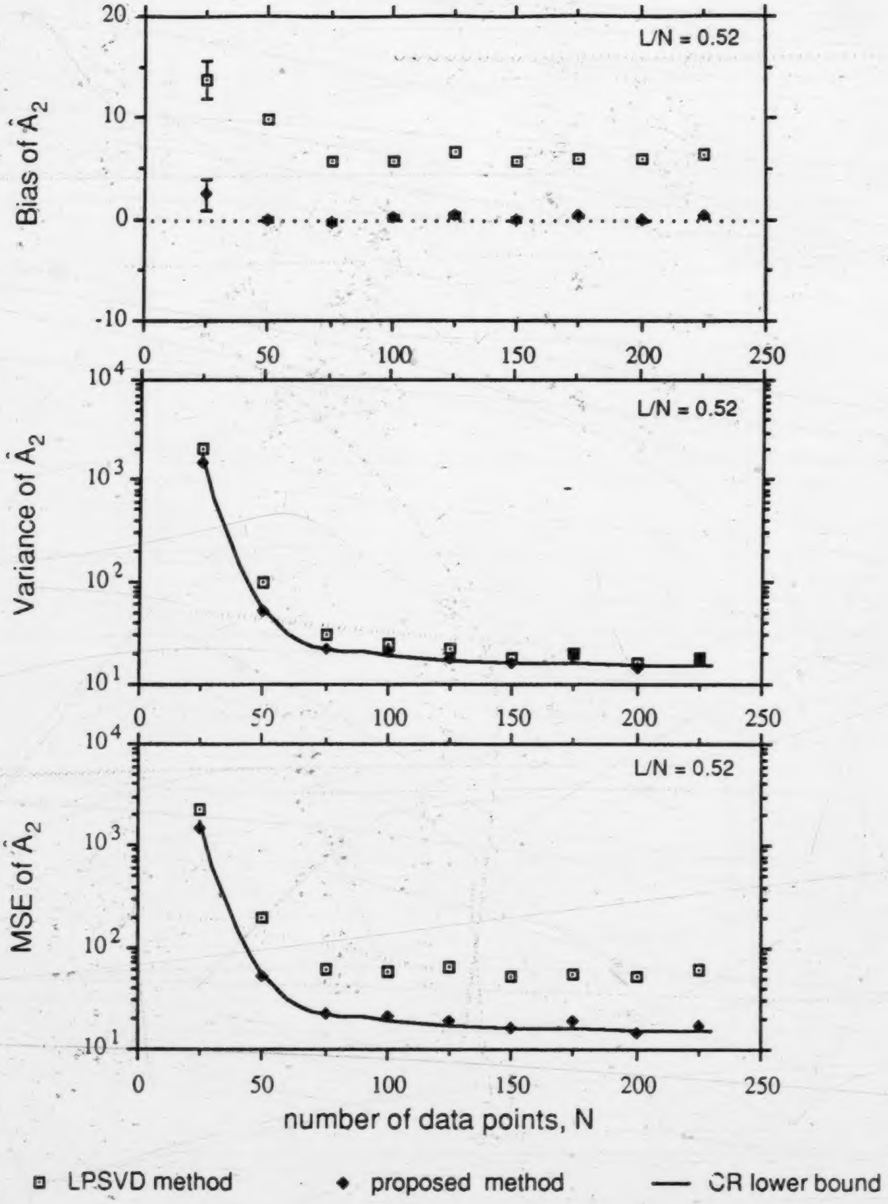


Figure 5.2.3.3 *Time domain amplitude of the 2-nd peak versus number of data points N. For other information, see Figure 5.2.3.1.*

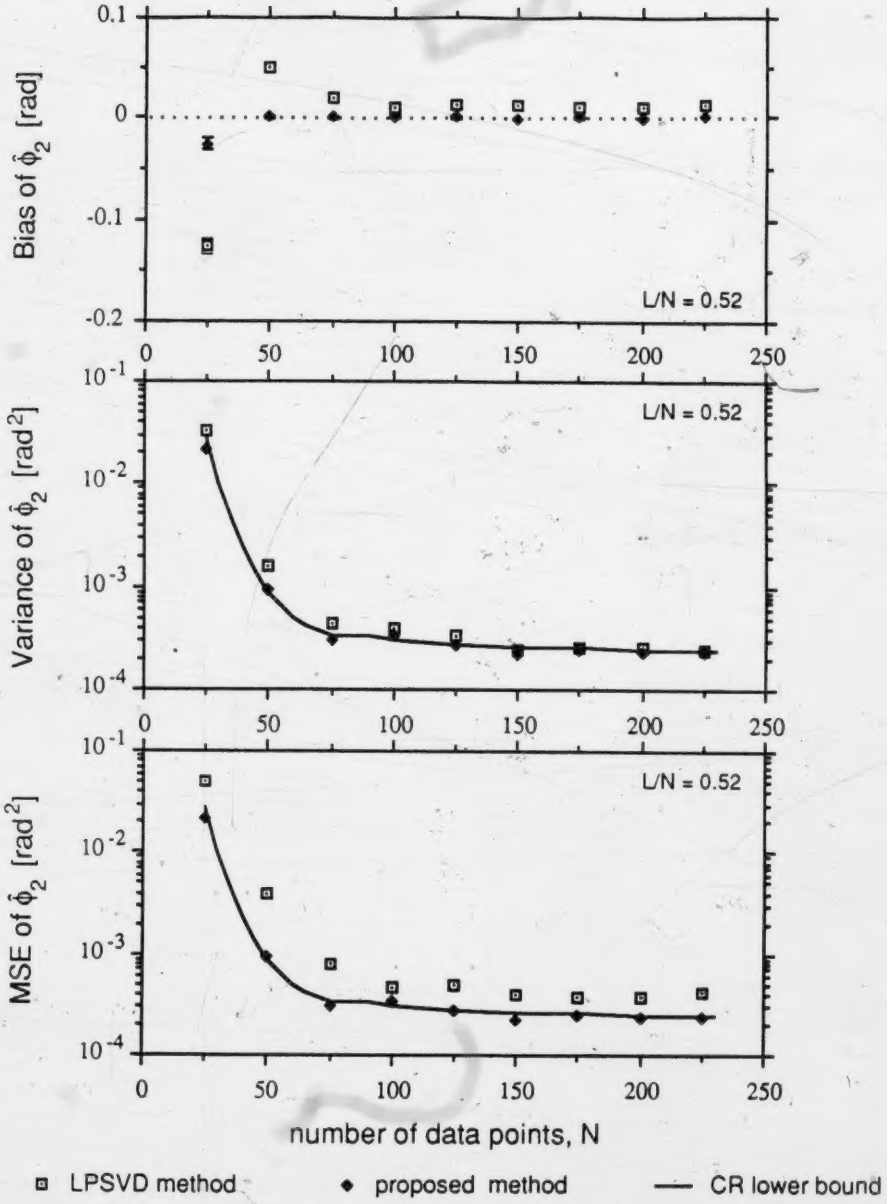


Figure 5.2.3.4 *Phase of the 2-nd peak versus number of data points N. For other information, see Figure 5.2.3.1.*

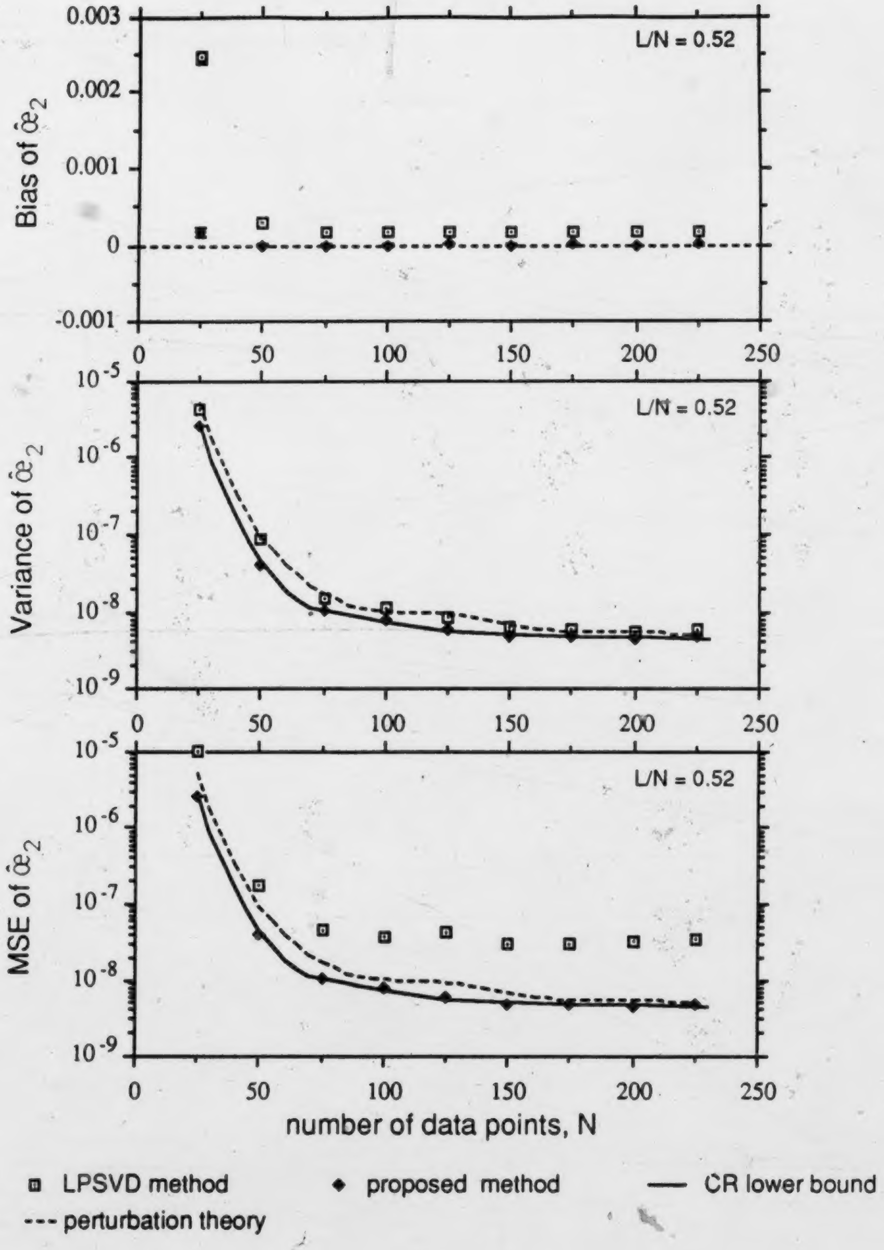


Figure 5.2.3.5 Normalized damping rate of the 2-nd peak versus number of data points N. For other information, see Figure 5.2.3.1.

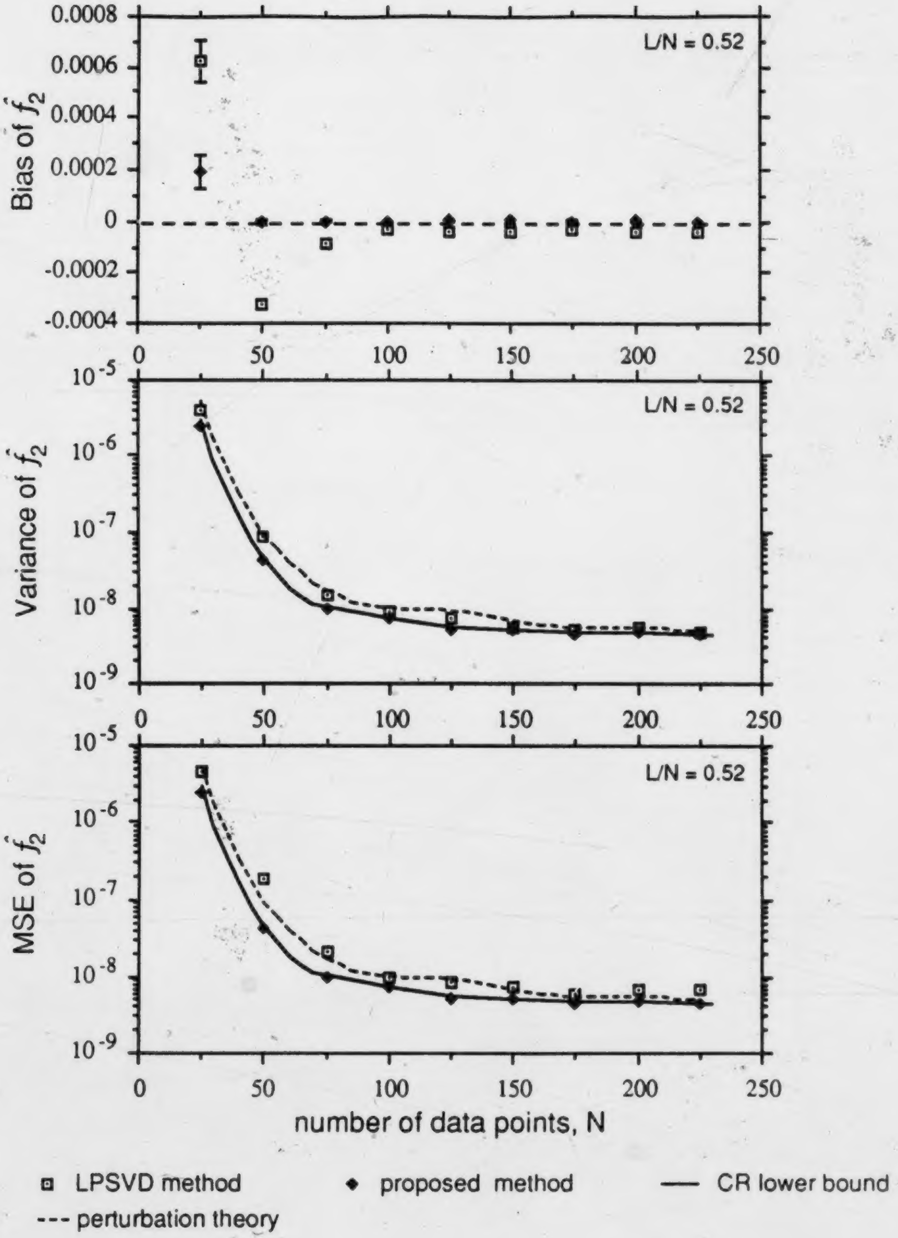


Figure 5.2.3.6 Normalized frequency of the 2-nd peak versus number of data points N. For other information, see Figure 5.2.3.1.

5.3 Indistinguishable signal and noise singular values

Recall that in section 4.2, the rank of A_B defined in Equation (4.2.1), where the y(i)'s were replaced by the s(i)'s as defined in Equation (1.1.1), was equal to the number of exponential components in an FID. But the presence of noise can perturb the singular values as demonstrated in section 3.4. Sometimes noise can be strong enough to obscure the signal singular values. In this section, we will examine a case where noise is strong enough to make one signal singular value indistinguishable from noise singular values.

The signal we are creating is one of those studied in the second example of section 3.4. The FID consisted of three components, sampled at 2500 Hz. The amplitudes were 250, 200 and 200, respectively. The phase was 0 for all components. The normalized damping rates were 0.00318, 0.0159 and 0.0134 (or equivalently, 50, 250 and 210 radian/second), respectively. The normalized frequencies were -0.3, -0.28 and -0.25, respectively. One hundred and fifty points were used by both the LPSVD and the proposed method. The tentative model order L was 75. Six hundred independent sequences of white Gaussian noise, with independent real and imaginary parts of variance $\sigma_w^2 = (15)^2 = 225$, were added to the three noise-free signals.

Figure 5.3.1 shows a representative spectrum of the signal obtained from 512 points DFT. Figure 5.3.2 shows the statistical result of the first eight singular values from the LPSVD method. Here, we see that the third singular value is not drastically different from the other smaller singular values. By looking only at the graph, one might mistakenly think that there are only two components, because the largest two singular values are well separated from the other singular values. Let us see what will result from the LPSVD method if we select the cut-off number of singular values $\hat{M}$ defined in Equation (4.2.3) as 2, 3 and 4.

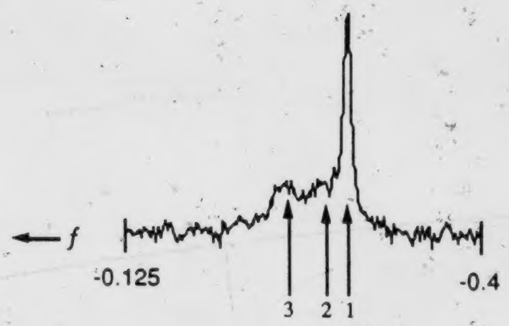


Figure 5.3.1 A representative spectrum obtained from 512 points for section 5.3 consisting of three peaks with time domain amplitudes 250, 200 and 200, normalized damping rates 0.00318, 0.0159 and 0.0134 (or equivalently, 50, 250, 210 rad/sec), normalized frequencies -0.3, -0.28 and -0.25, respectively. Time domain noise standard deviation σ_W is 15.

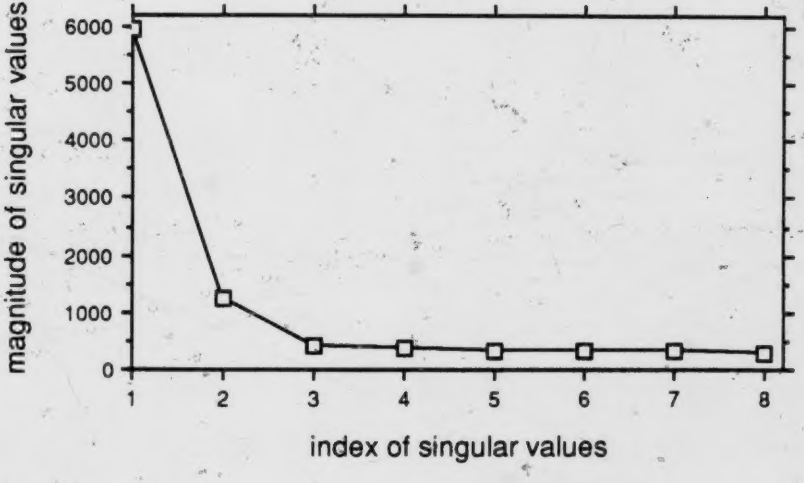


Figure 5.3.2 Statistical result of the magnitude of the first eight singular values of the LPSVD method. One hundred and fifty points are used and the tentative model order is 75. Time domain noise standard deviation σ_W is 15. For other information, see Figure 5.3.1

In Figure 5.3.3, we show the probability of finding a different number of peaks in the noise corrupted FID signals versus the cut-off number of singular values of the LPSVD method. If we select two principal singular values, the LPSVD method detects only two peaks almost 100% of the time. They are the two located at normalized frequencies of -0.3 and -0.25. Selecting a cut-off of three singular values, we detect two components 20% of the time, three components 73% of the time, and 7% of the time more than three components. Choosing four cut-off singular values, we have a 7% chance of finding only two components, a 54% chance of three components, and a 39% chance of more than three components. Increasing the cut-off number of singular values results in a greater chance of detecting extraneous peaks, and this is consistent with Kumaresan and Tufts' observation (1982). On balance, it might be better to select the cut-off number of singular values, $\hat{M}$, at least one more than the expected number of peaks, so that the chance of picking less than the expected number of peaks is small. However, we might pay the price of finding spurious peaks.

Figures 5.3.4 to 5.3.17 show the statistical results for the performance measures of the spectral parameter estimates from the LSVD method and the proposed method for $\hat{M}$ equal to three and four. Here, we see that increasing the cut-off number of singular values generally decreases the biases and variances of the estimates. Unfortunately, the estimates from both methods are still far away from the theoretical lowest CR bounds.

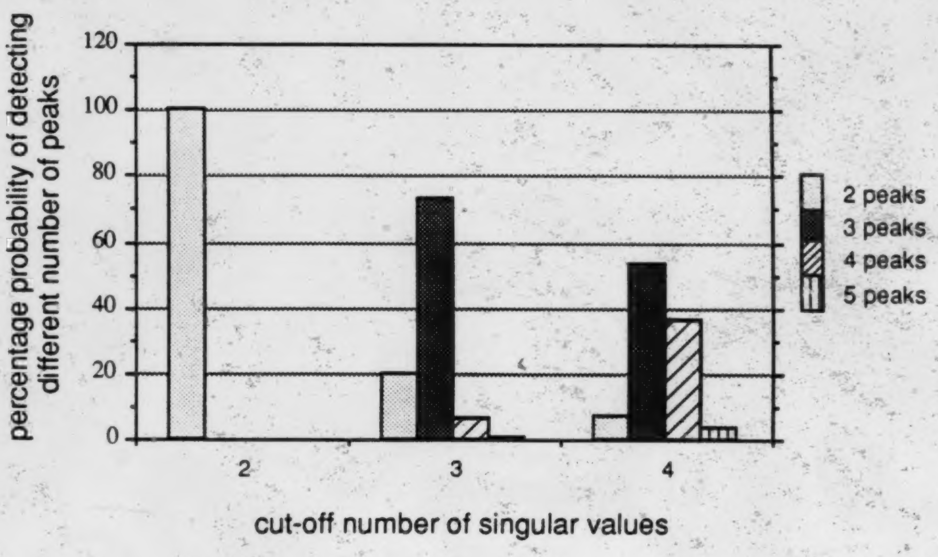


Figure 5.3.3 *Percentage probability of finding different number of peaks versus the cut-off number of singular values of the LPSVD method for the example in section 5.3. For other information, see Figure 5.3.2.*

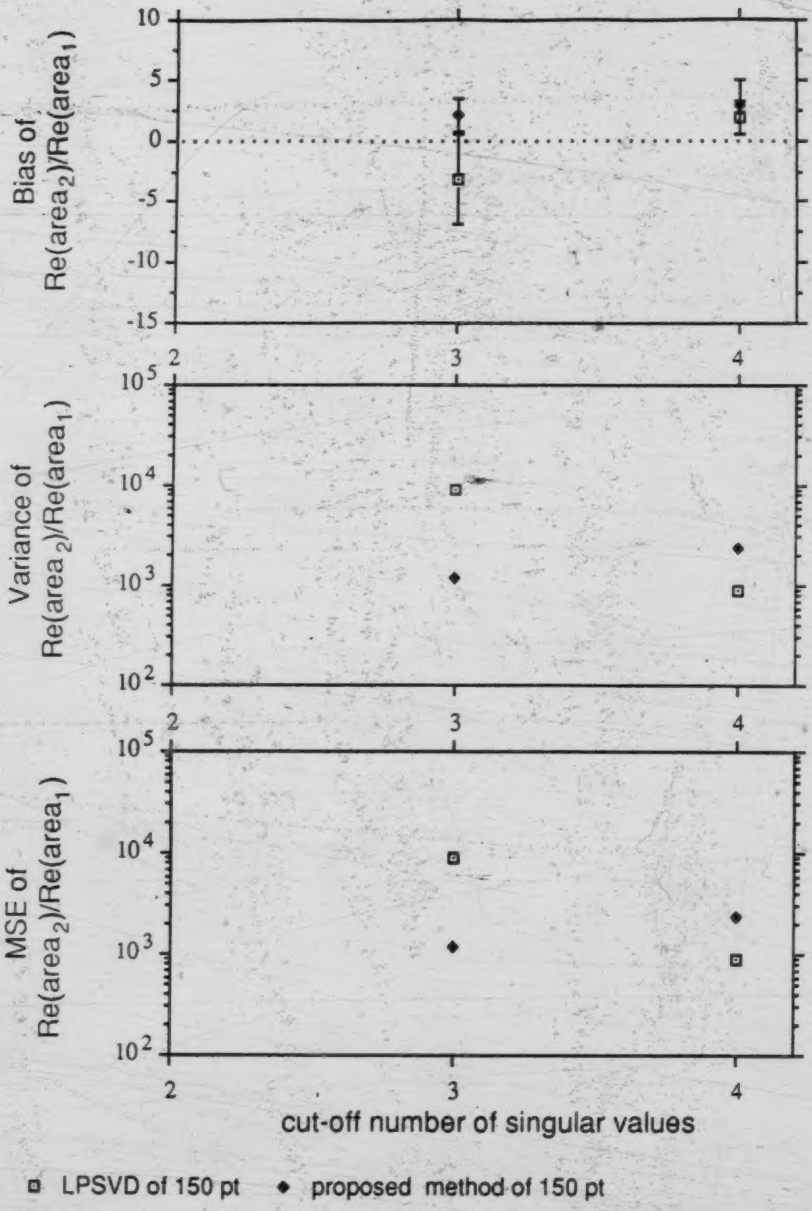


***Figure 5.3.4** The area ratio of the 2-nd peak to the 1-st peak from the real spectrum versus the cut-off number of singular values of the LPSVD method. The squares and diamonds, for the LPSVD method and the proposed method, respectively, are obtained from over 600-run simulations. Bias and MSE are defined in section 5.1. For other information, see Figure 5.3.2.*

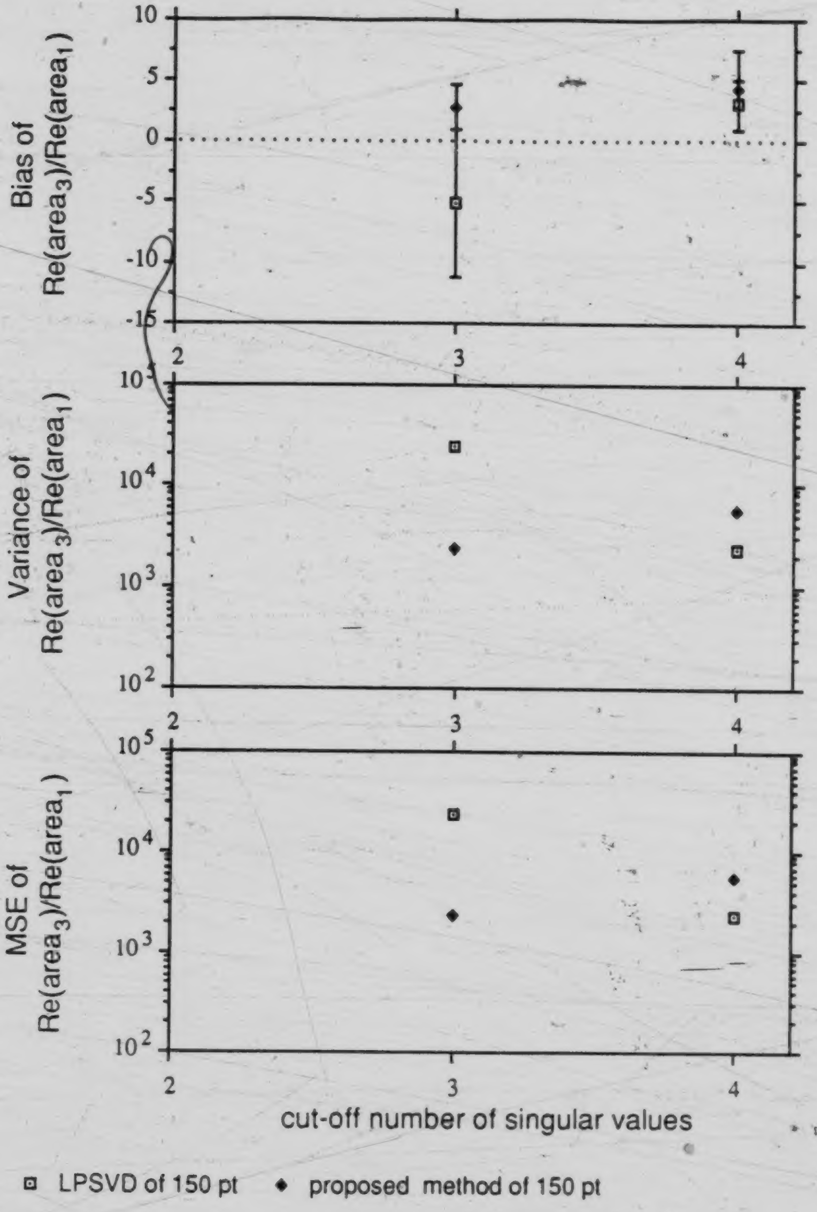


Figure 5.3.5 The area ratio of the 3-rd peak to the 1-st peak from the real spectrum versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.

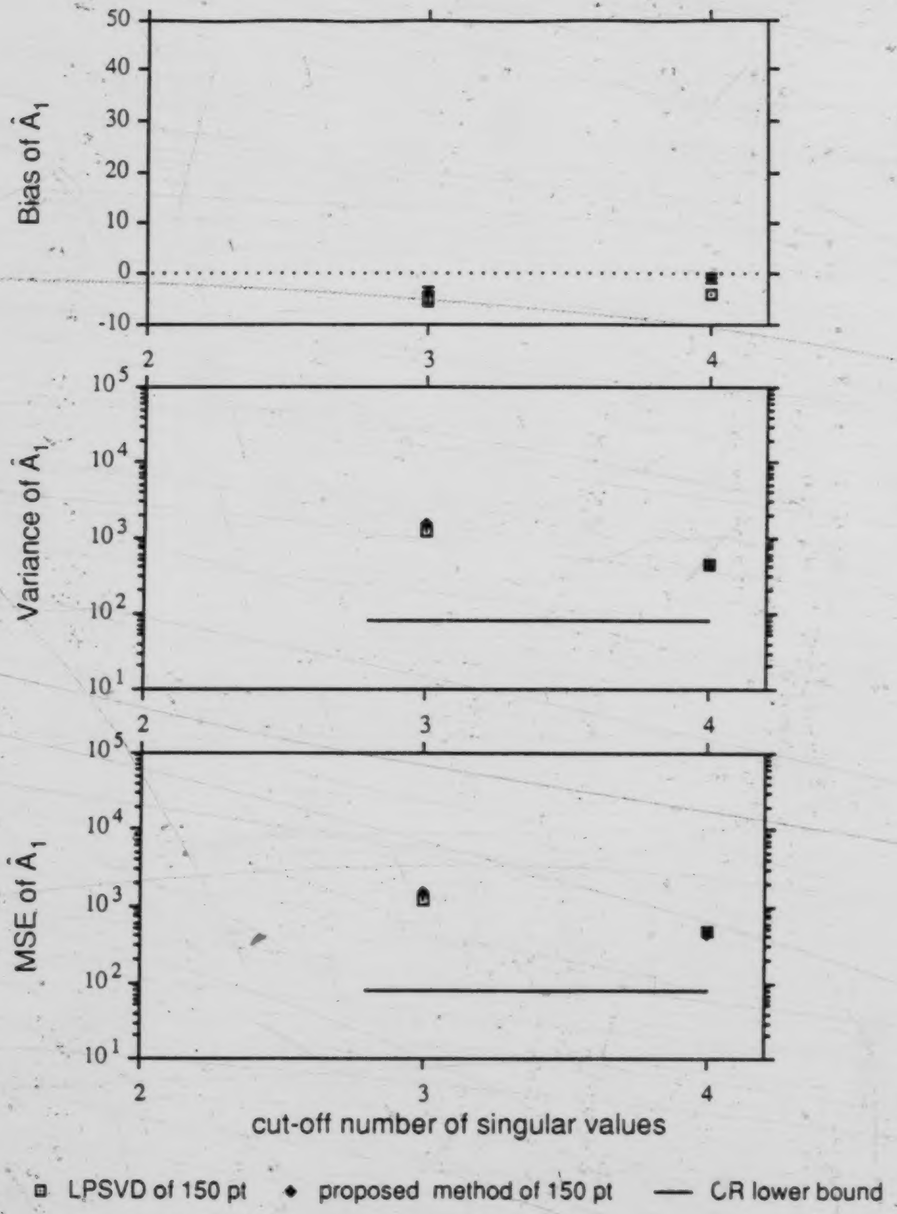


Figure 5.3.6 *Time domain amplitude of the 1-st peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.*

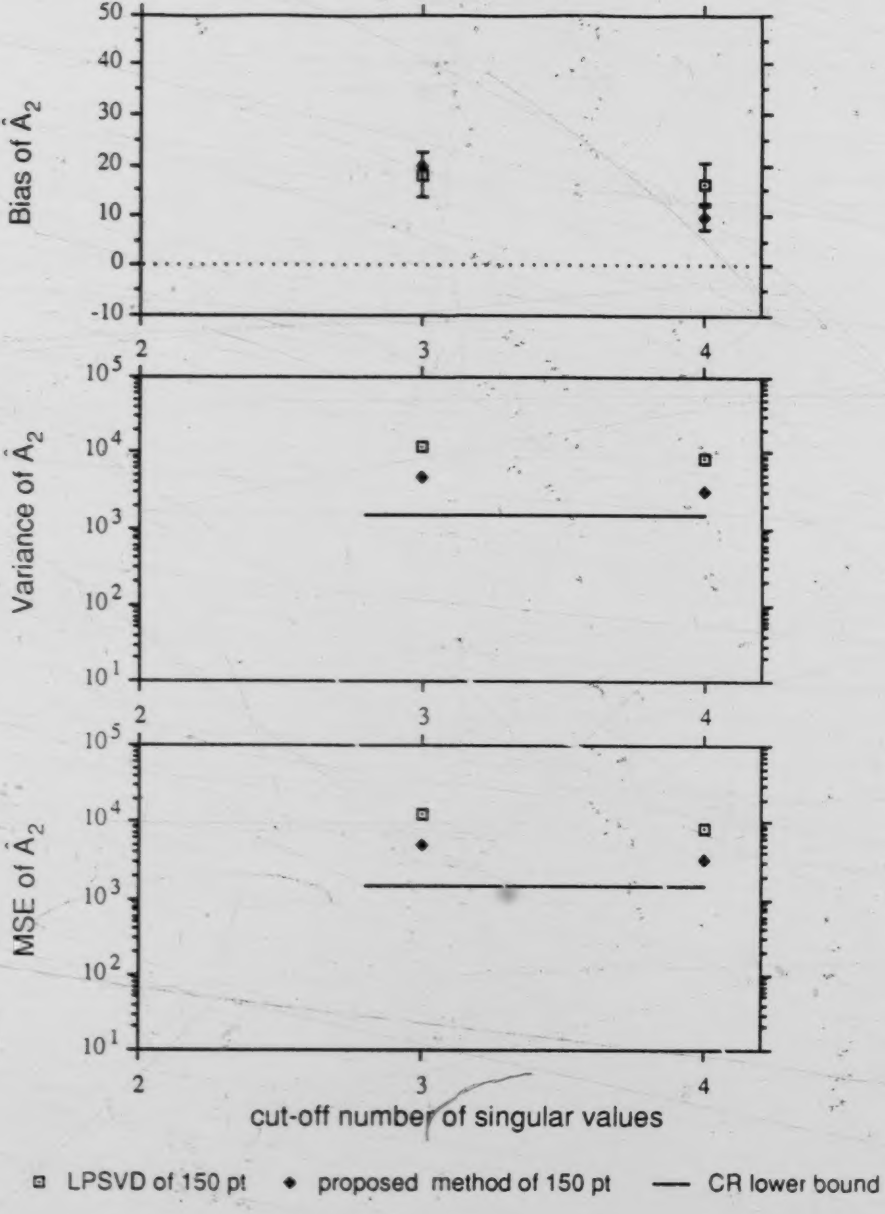


Figure 5.3.7 Time domain amplitude of the 2-nd peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.

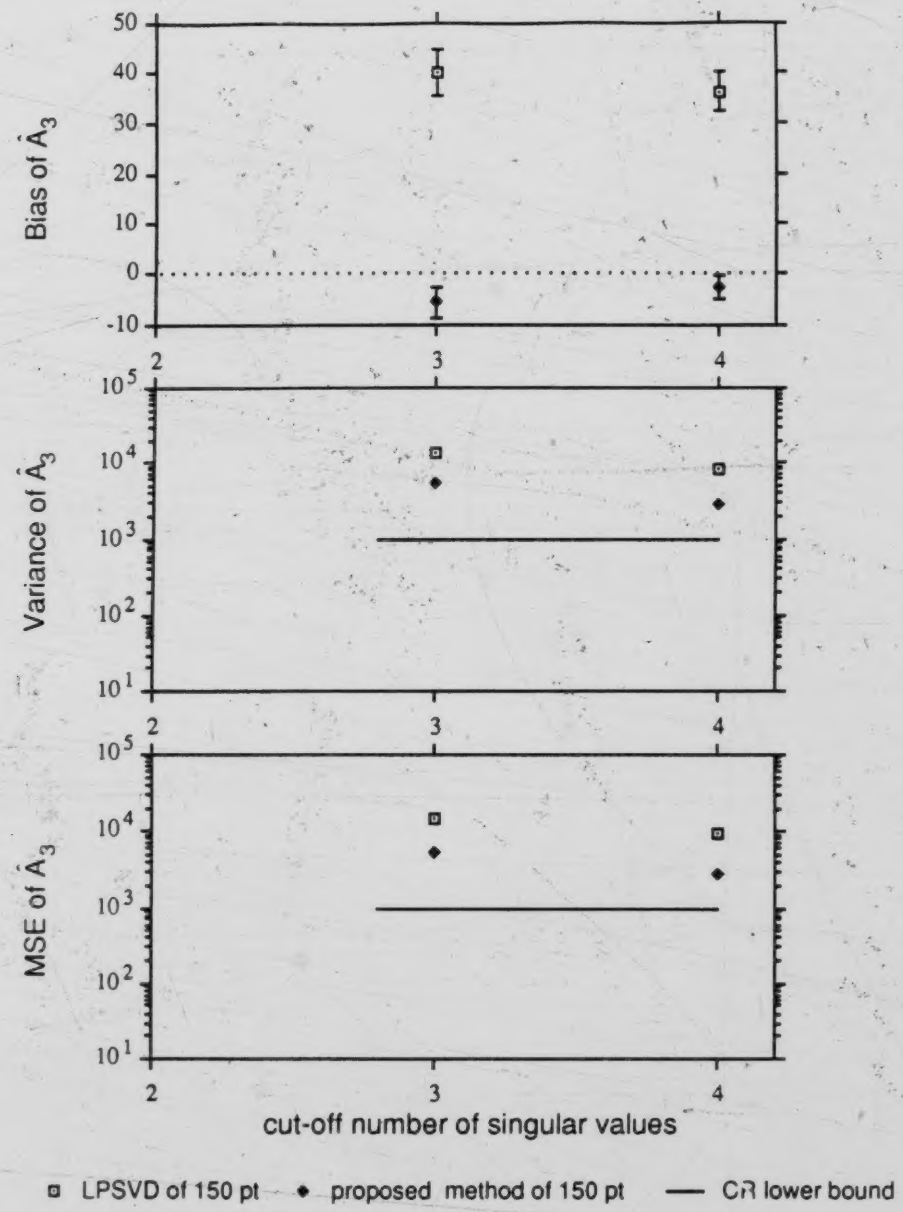


***Figure 5.3.8** Time domain amplitude of the 3-rd peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.*

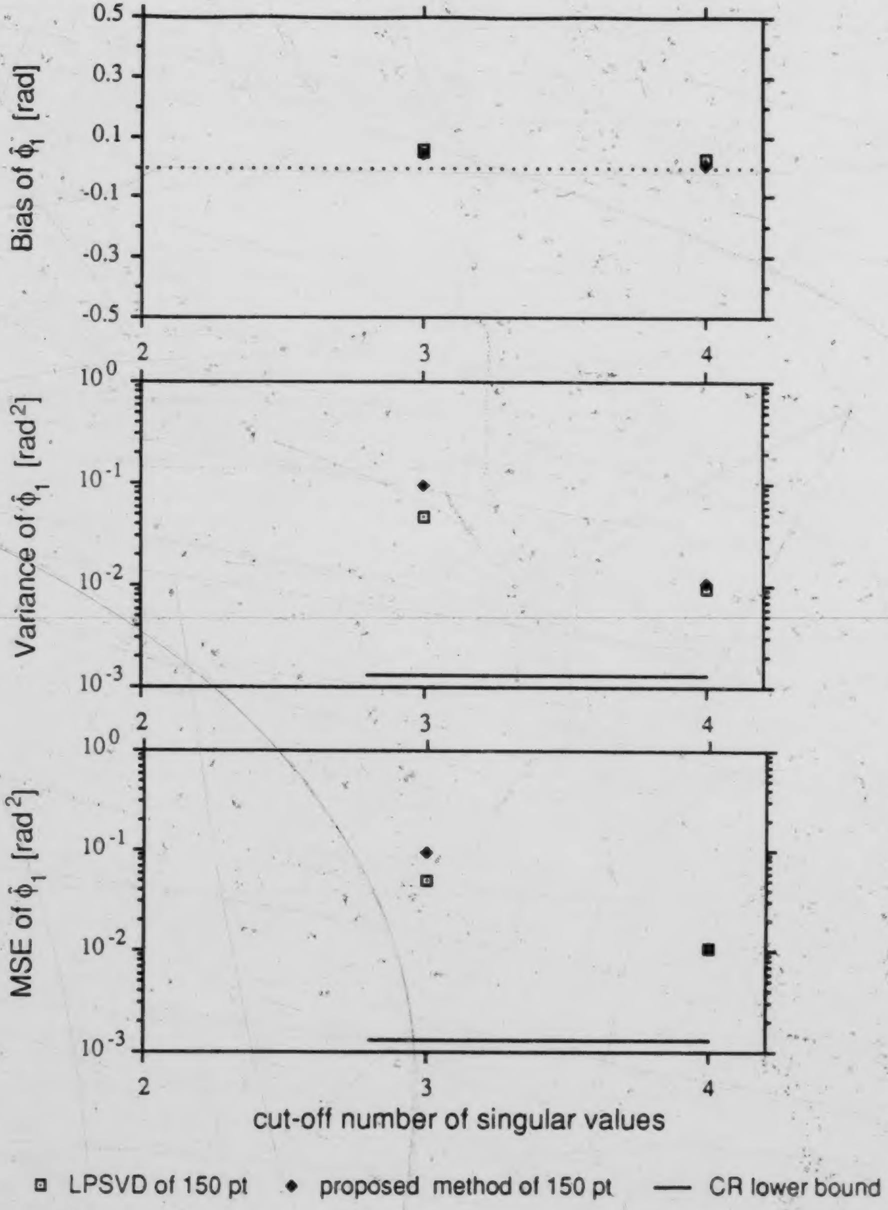


Figure 5.3.9 *Phase of the 1-st peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.*

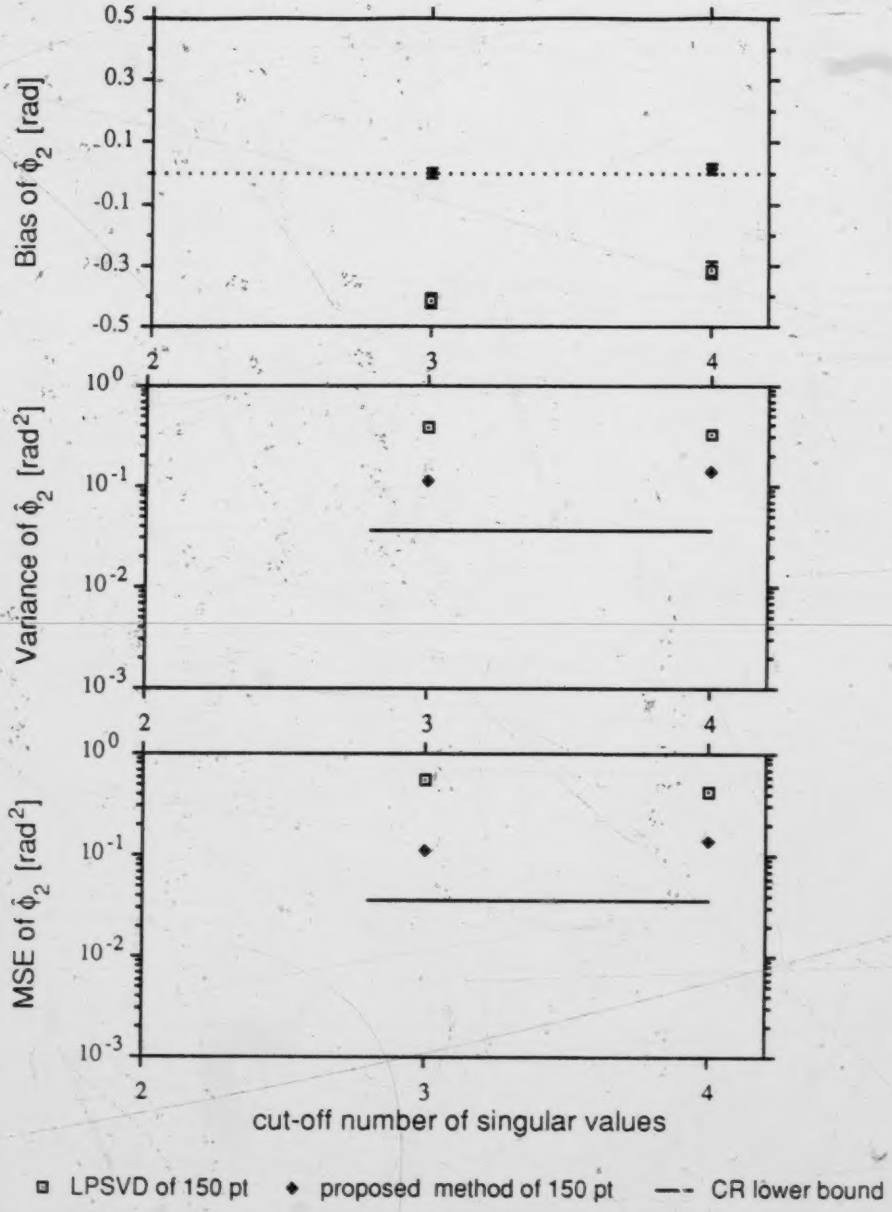


***Figure 5.3.10** Phase of the 2-nd peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.*

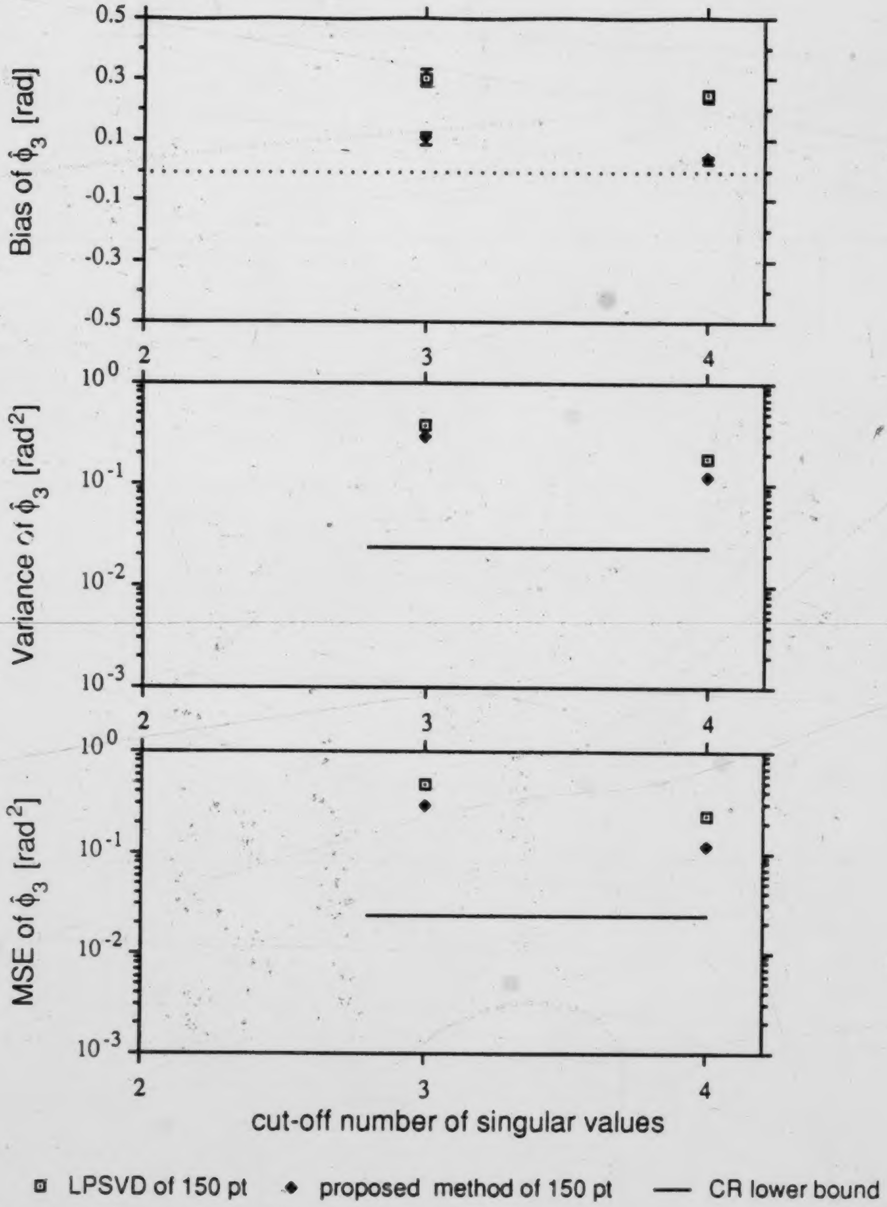


***Figure 5.3.11** Phase of the 3-rd peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.*

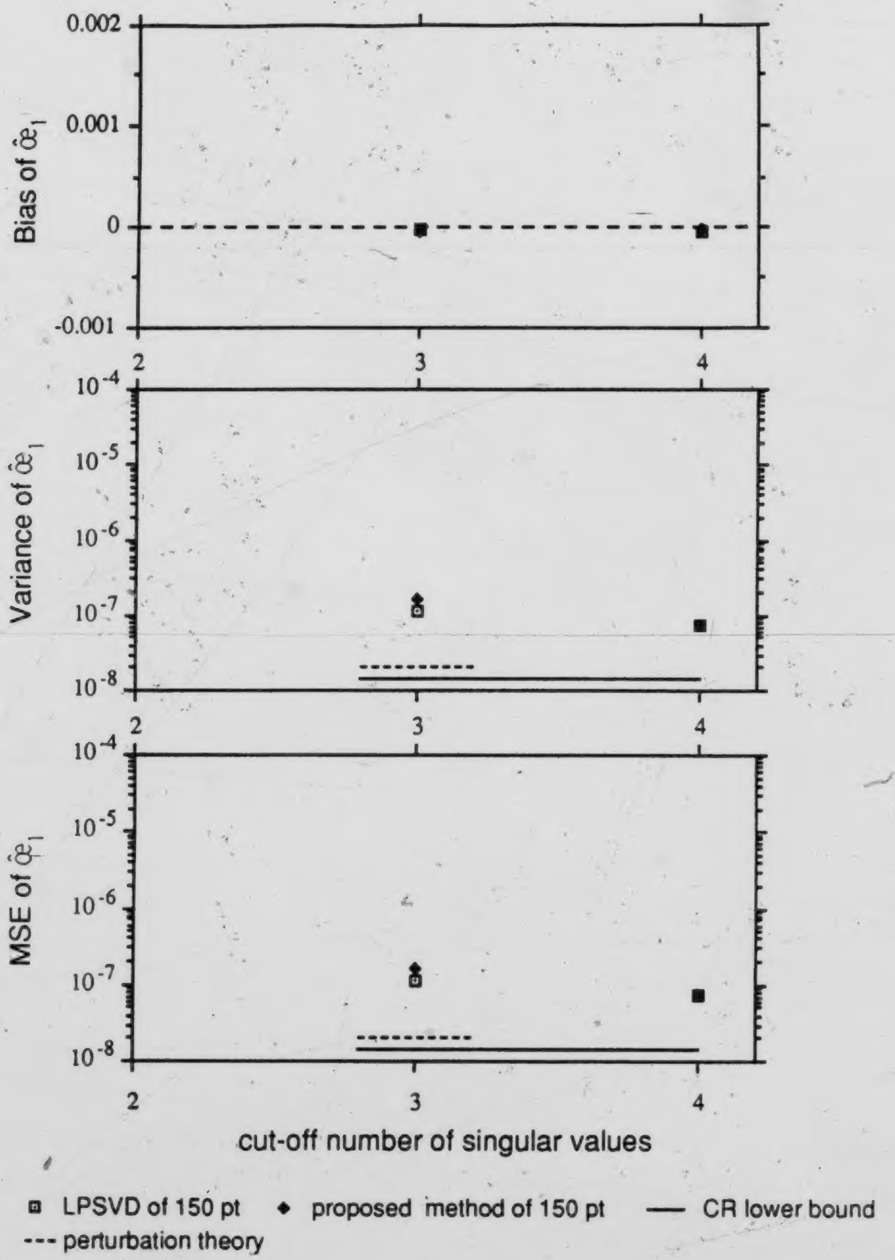


Figure 5.3.12 Normalized damping rate of the 1-st peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.

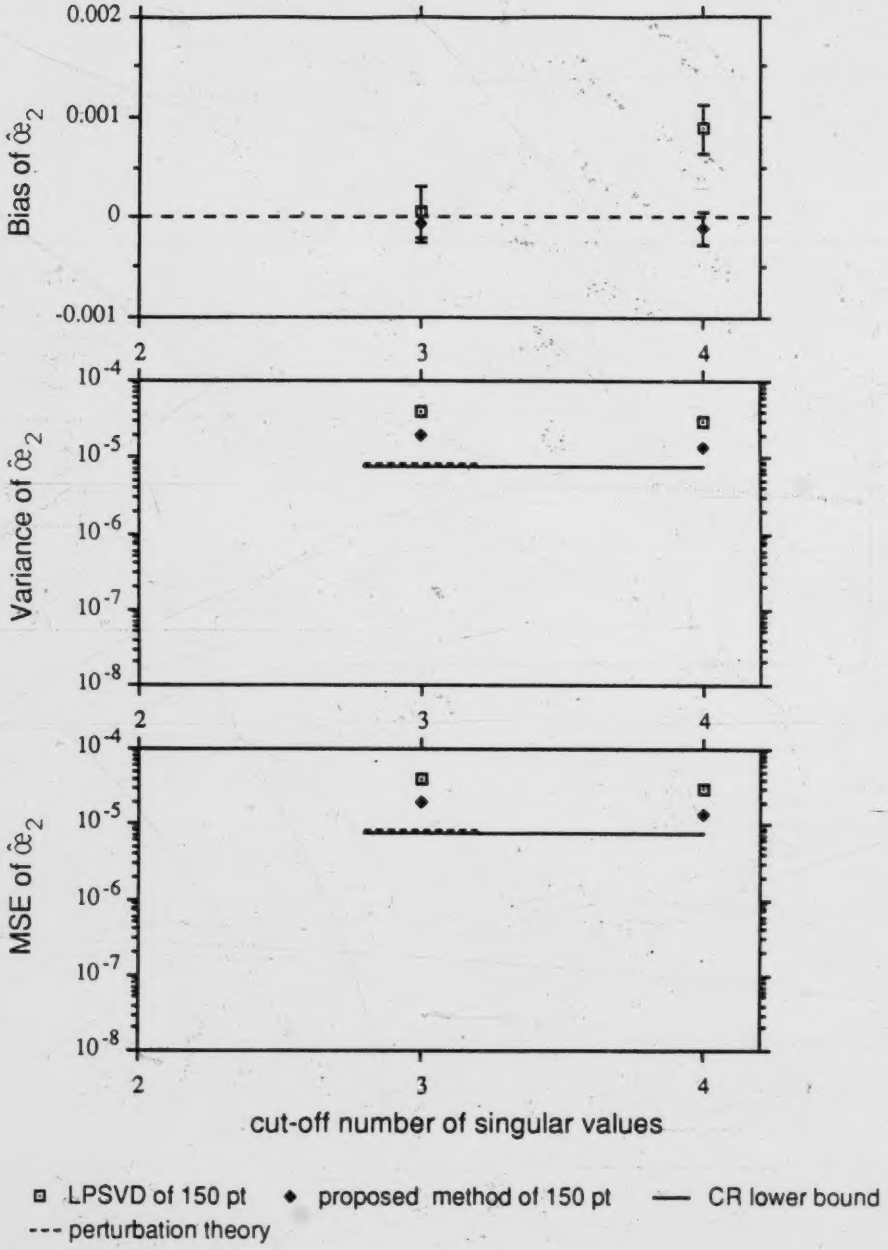


Figure 5.3.13 Normalized damping rate of the 2-nd peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.

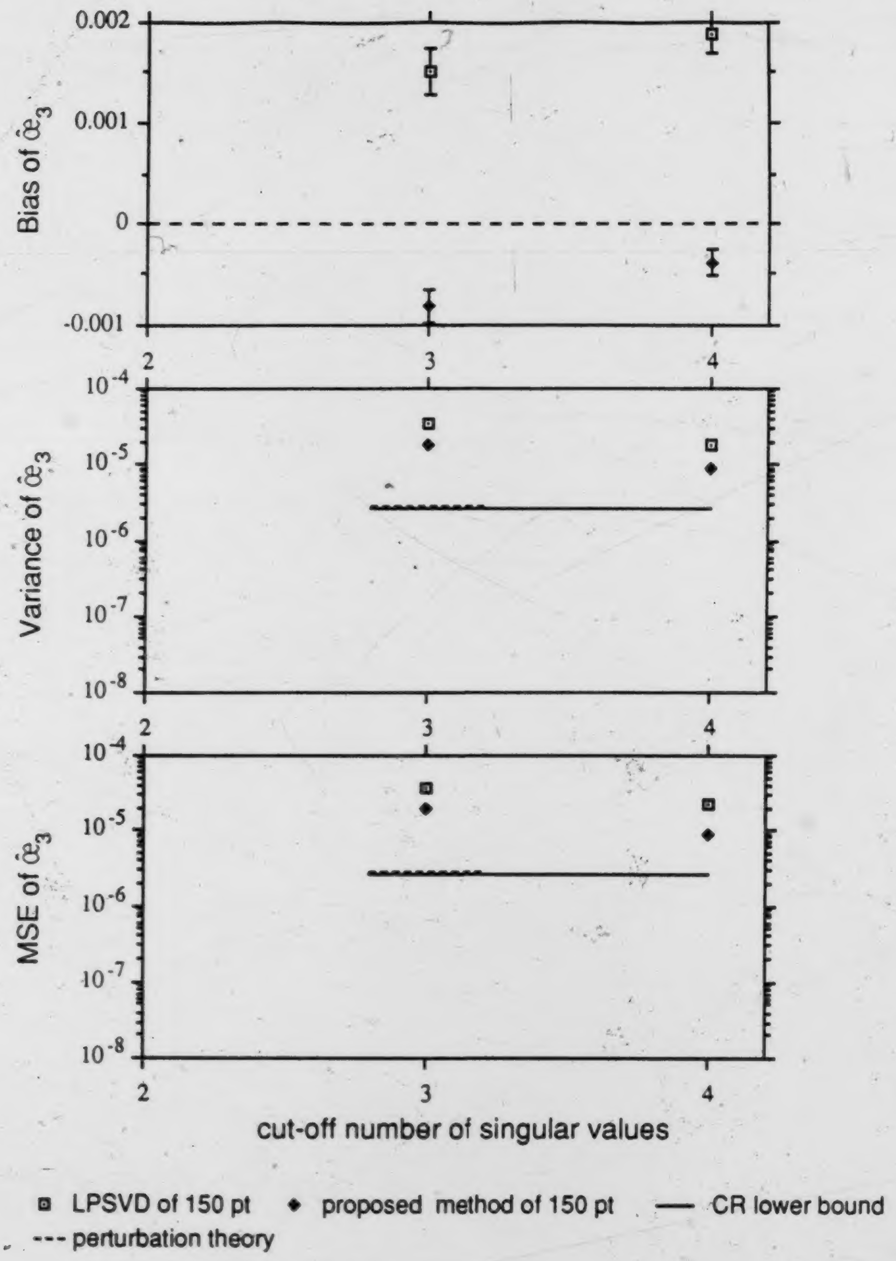


***Figure 5.3.14** Normalized damping rate of the 3-rd peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.*

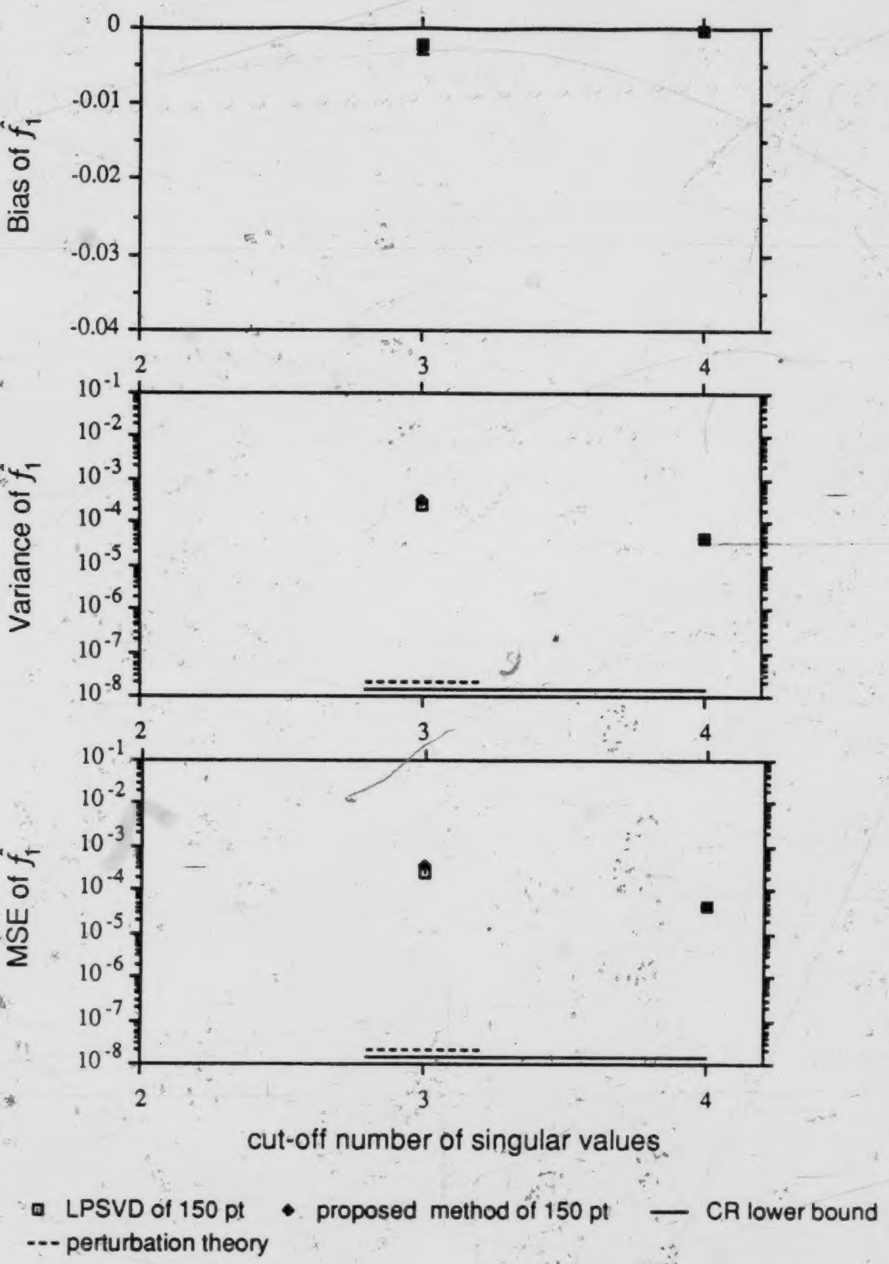


Figure 5.3.15 Normalized frequency of the 1-st peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.

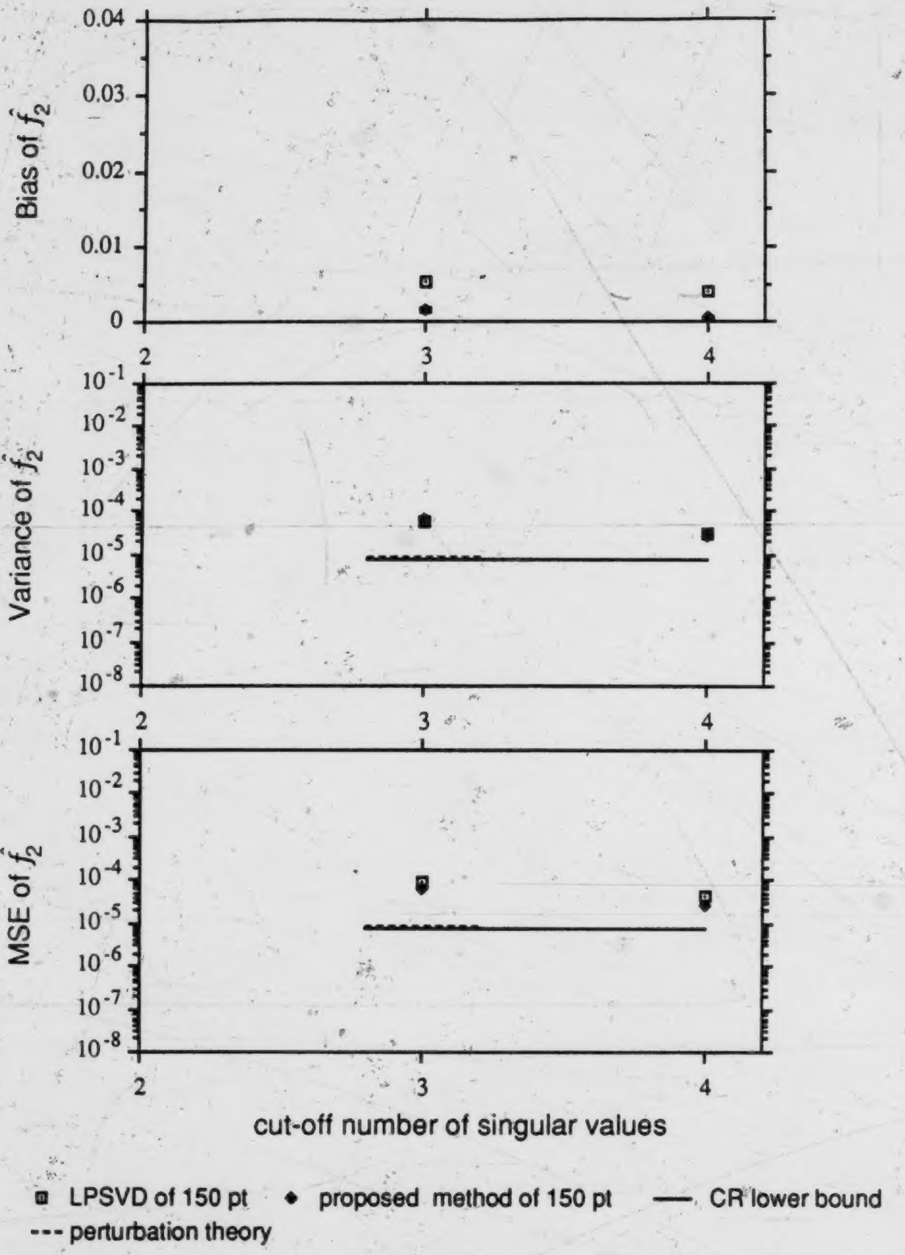


Figure 5.3.16 *Normalized frequency of the 2-nd peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.*

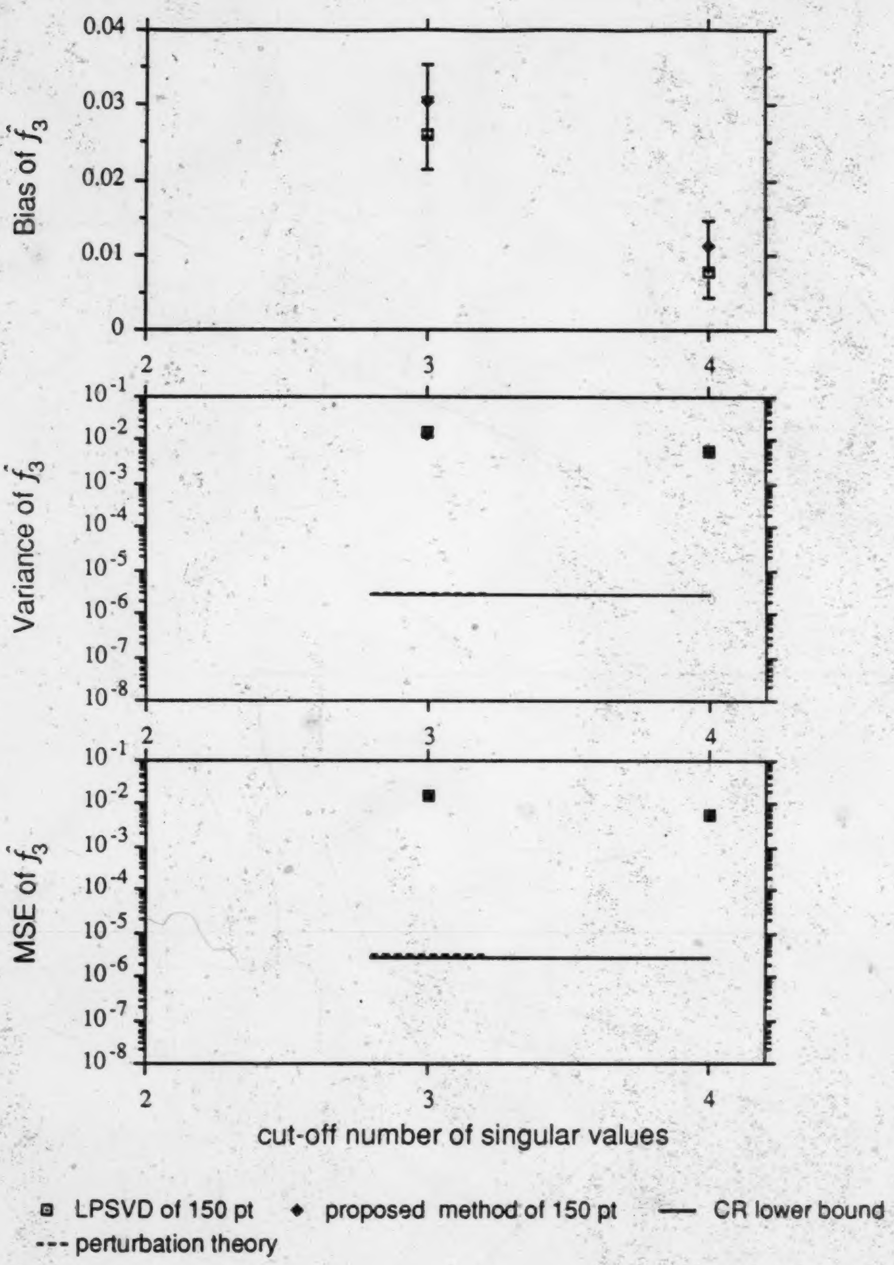


Figure 5.3.17 Normalized frequency of the 3-rd peak versus the cut-off number of singular values of the LPSVD method. For other information, see Figure 5.3.2.

5.4 Effect of noise (i.e., spectral sensitivity)

In this section, we will examine the effect of noise on the three methods: the LPSVD method, the proposed method and the DFT method. Recall that the noise we will consider is zero mean complex white Gaussian noise with uncorrelated real and imaginary parts each having a variance of $\sigma_w{}^2$.

If we define the signal-to-noise ratio of the i-th component in the time domain $(SNR_i)_t$ to be the ratio of the average power of the noise-free signal to the average power of the noise, then one can show that for exponentially decaying signals as defined in Equation (1.1.1), we have

$$(SNR_i)_t \quad = \quad \frac{1}{2N} \, \frac{1}{2\pi \, \alpha_i \, \Delta t} \, \frac{A_i^2}{2\sigma_w^2} \, , \qquad (5.4.1)$$

where A_i is the amplitude of the i-th exponentially decaying component in the signal defined in Equation (1.1.1), α_i is the corresponding damping factor in frequency unit, N is the total number of time domain data points; Δt is the sampling interval in time domain, and $2\sigma_w{}^2$ is the time domain variance of the white complex Gaussian noise. The derivation of Equation (5.4.1) is given in Appendix 19.

Since NMR spectroscopists are more familiar with a SNR defined in the frequency domain, we have derived, in Appendix 20, the statistical relationship between white complex Gaussian noise of zero mean and with variance $2\sigma_w{}^2$ in the time domain, and the corresponding Gaussian noise in the frequency domain. It turns out that the relationship is as follows:

$$(\sigma_w)_f \quad = \quad \Delta t \, \sigma_w \sqrt{N} \, , \qquad (5.4.2)$$

where $(\sigma_w)_f$ is the noise standard deviation in the frequency domain.

The SNR of the i-th peak in the frequency domain (denoted by $(SNR_i)_f$) is usually defined as

$$(SNR_i)_f \equiv \frac{\text{i-th peak-height}}{1.96*(\text{RMS noise in frequency domain})} , \qquad (5.4.3)$$

where the RMS (root-mean-square) value of the noise defined as

$$\text{RMS} = \sqrt{\frac{1}{N-1} \sum_{n=0}^{N-1} \left[w_f(n) - \frac{1}{N} \sum_{m=0}^{N-1} w_f(m) \right]^2} \qquad (5.4.4)$$

is an unbiased estimate of the noise standard deviation in the frequency domain. $w_f(n)$ is the noise value in the frequency domain. The factor 1.96 in Equation (5.4.3) is the ratio between the peak value (not peak-to-peak) to the RMS value of the Gaussian noise, for a confidence level of 95% (Barford, 1985, p. 148).

The relationship of Lorentzian peak-height and the corresponding time domain parameters, given in Equation (A18.7), is

$$\text{i-th peak-height} = \frac{A_i}{2\pi\alpha_i} . \qquad (5.4.5)$$

Therefore, the relationship between the SNR of the i-th Lorentzian peak in the frequency domain in white complex Gaussian noise and the corresponding time domain parameters is

$$(SNR_i)_f = \frac{A_i}{2\pi\alpha_i} \frac{1}{1.96\,(\sigma_w)_f} , \qquad (5.4.6)$$

$$= \frac{1}{1.96\ \sqrt{N}\ 2\pi\alpha_i\,\Delta t} \frac{A_i}{\sigma_w} . \qquad (5.4.7)$$

Equation (5.4.6) is obtained from Equation (5.4.3) using Equation (5.4.5). Equation (5.4.7) is obtained by substituting Equation (5.4.2) into Equation (5.4.6).

Now we will present an example to see the effect of noise.

Example 5.4.1 :

In this example we will use the same signal parameters as in Example 5.2.1 except that N =150 and L = 60 for the LPSVD method and for the proposed method. For the DFT method, we use three different values of N. They are 128, the multiple power of 2 that is nearest to 150; 512 and 1024 which are often acquired in NMR FID. Different amounts of complex white Gaussian noises ranging from $\sigma_w = 2.5$ to 125 are added to the noise-free signals, and over six hundred independent runs are used for each choice of σ_w. Since $\Delta t = 4 \times 10^{-4}$ sec, this range of noise is equivalent from 833.33 to 0.33 for $(SNR)_t$ with N = 150. In the frequency domain with N = 128, 512 and 1024, this range of noise gives us signal-to-noise ratios $(SNR)_f$ from 225.48 to 4.51, 112.74 to 2.25, and 79.72 to 1.59, respectively.

Figure 5.4.1.1 shows some representative spectra obtained using DFT, where the different SNRs arise from the different values of σ_w. Figure 5.4.1.2 shows the statistical result of the first five singular values at different SNRs. As we see, for this particular signal embedded in this range of noise, no ambiguity arises in the correct number of components. Figure 5.4.1.3 shows us how noise perturbs the singular values. Basically, all singular values are affected, but the noise related singular values vary linearly with the time domain noise standard deviation σ_w, and each noise singular value has a different rate of change of magnitude.

Figures 5.4.1.4 to 5.4.1.12 show the results of the performance measures of the estimates of spectral parameters from different methods. From these figures, we see that the variance estimates obtained from both the LPSVD and the proposed method vary linearly with noise variance. The MSE estimates obtained from the DFT method seem to vary linearly with noise variance at low SNRs, but

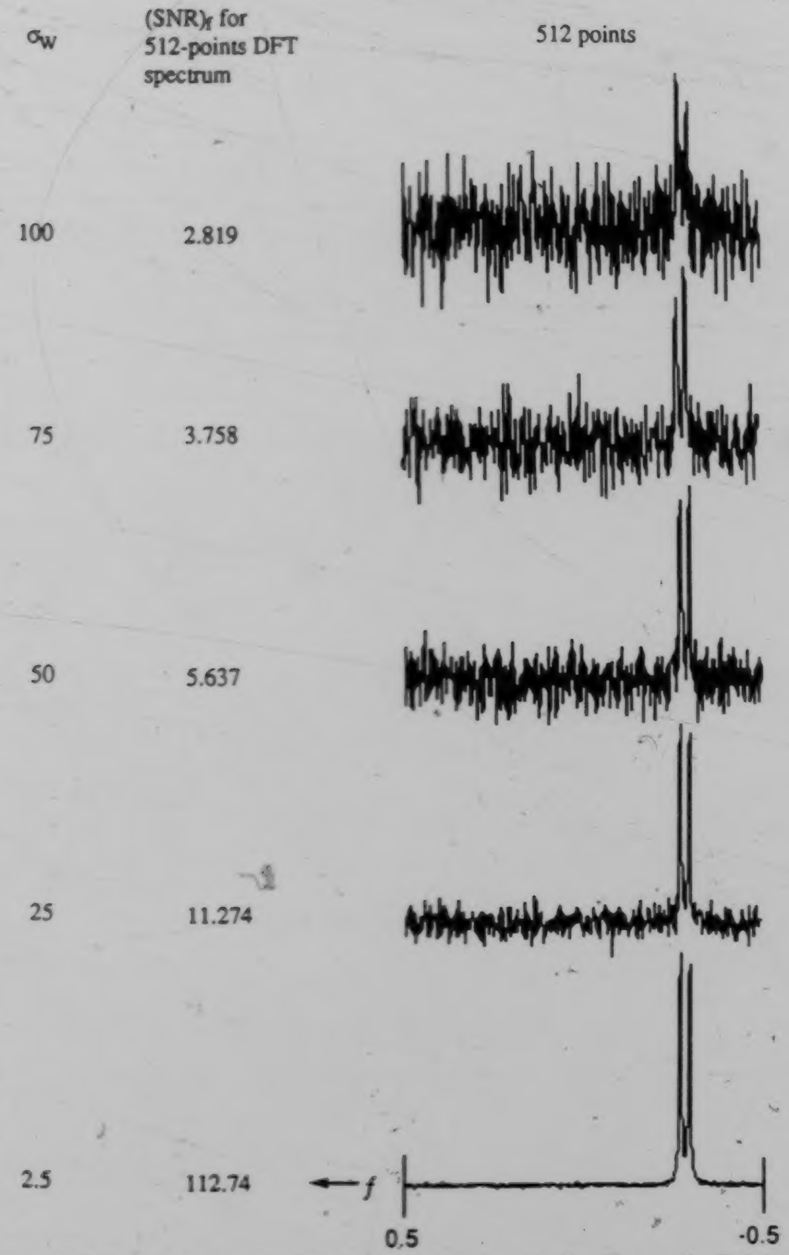


Figure 5.4.1.1 *Some representative DFT spectra for Example 5.4.1, consisting of two equal amplitude (250), equal normalized damping rate (0.00318) sinusoids of normalized frequency separation 0.025 at different SNRs. N is the number of data points;* σ_W *is the noise standard deviation in the time domain defined in Equation (1.1.1);* $(SNR)_t$ *is the time domain SNR defined in Equation (5.4.1);* $(SNR)_f$ *is the frequency domain SNR defined in Equation (5.4.7).*

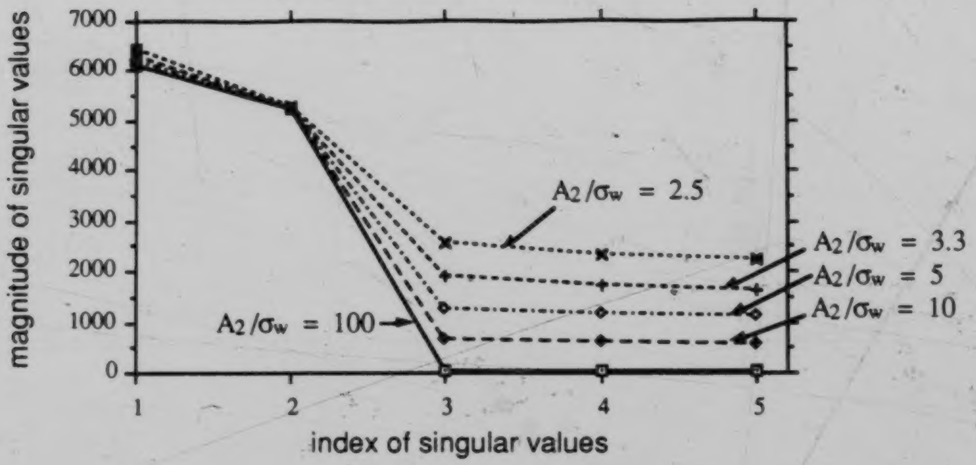


Figure 5.4.1.2 Statistical result of the first five singular values of the LPSVD method for the noise corrupted signals described in Example 5.4.1. σ_w is the time domain noise standard deviation ; A_2 (= 250) is the time domain amplitude of the 2-nd component of the noise-free signal. Number of data points used was 150. Tentative model order L was 60. For other information, see Figure 5.4.1.1.

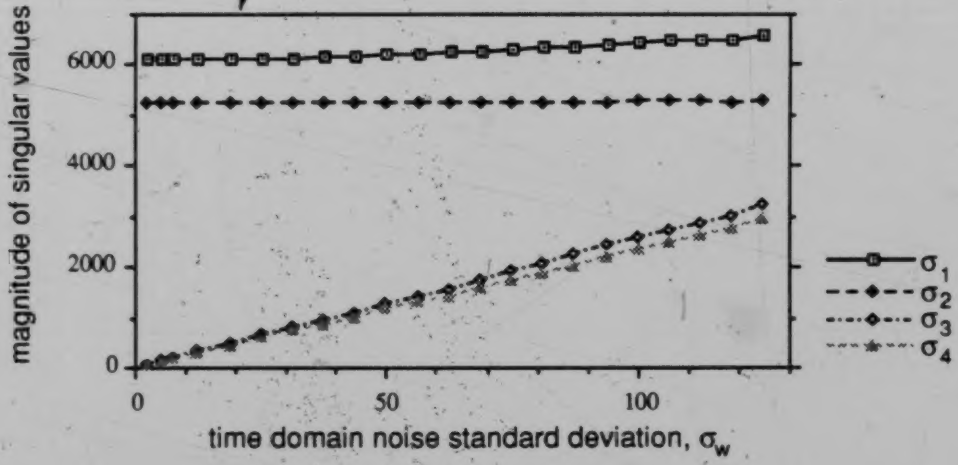


Figure 5.4.1.3 Magnitude of the first four singular values of the LPSVD method versus input time domain noise standard deviation σ_w. For other information, see Figure 5.4.1.2.

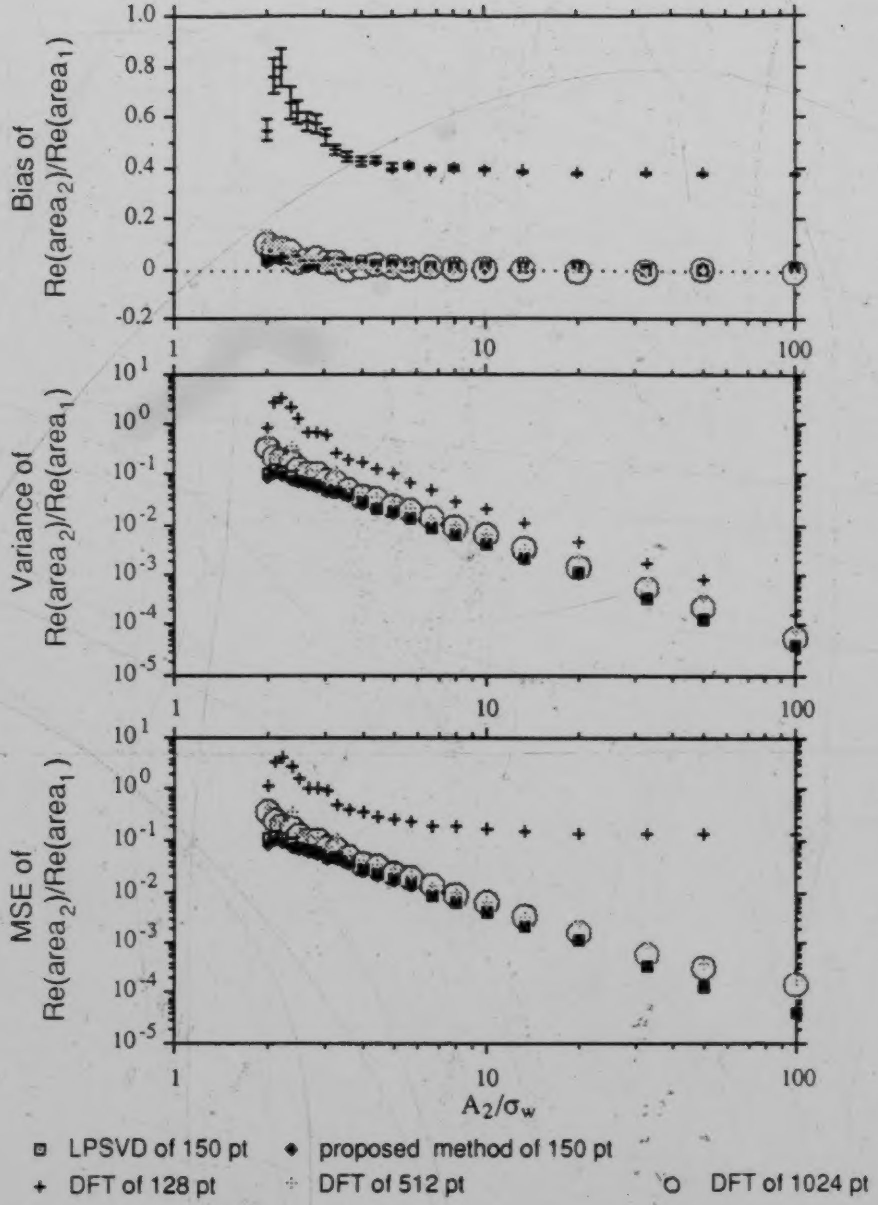


Figure 5.4.1.4 *The area ratio of the two peaks from the real spectrum versus* A_2/σ_w, *where* A_2 *is the time domain amplitude of the 2-nd peak and* σ_w *is time domain noise standard deviation.* $\Delta t = 4 \times 10^{-4}$ *sec;* $A_1 = A_2 = 250$; $\phi_1 = \phi_2 = 0$; $\alpha_1 = \alpha_2 = 0.00318$; $f_1 - f_2 = -0.3 - (-0.275) = 0.025$. *The tentative model order for the LPSVD method is 60. All estimates are obtained from over 600-run simulations.*

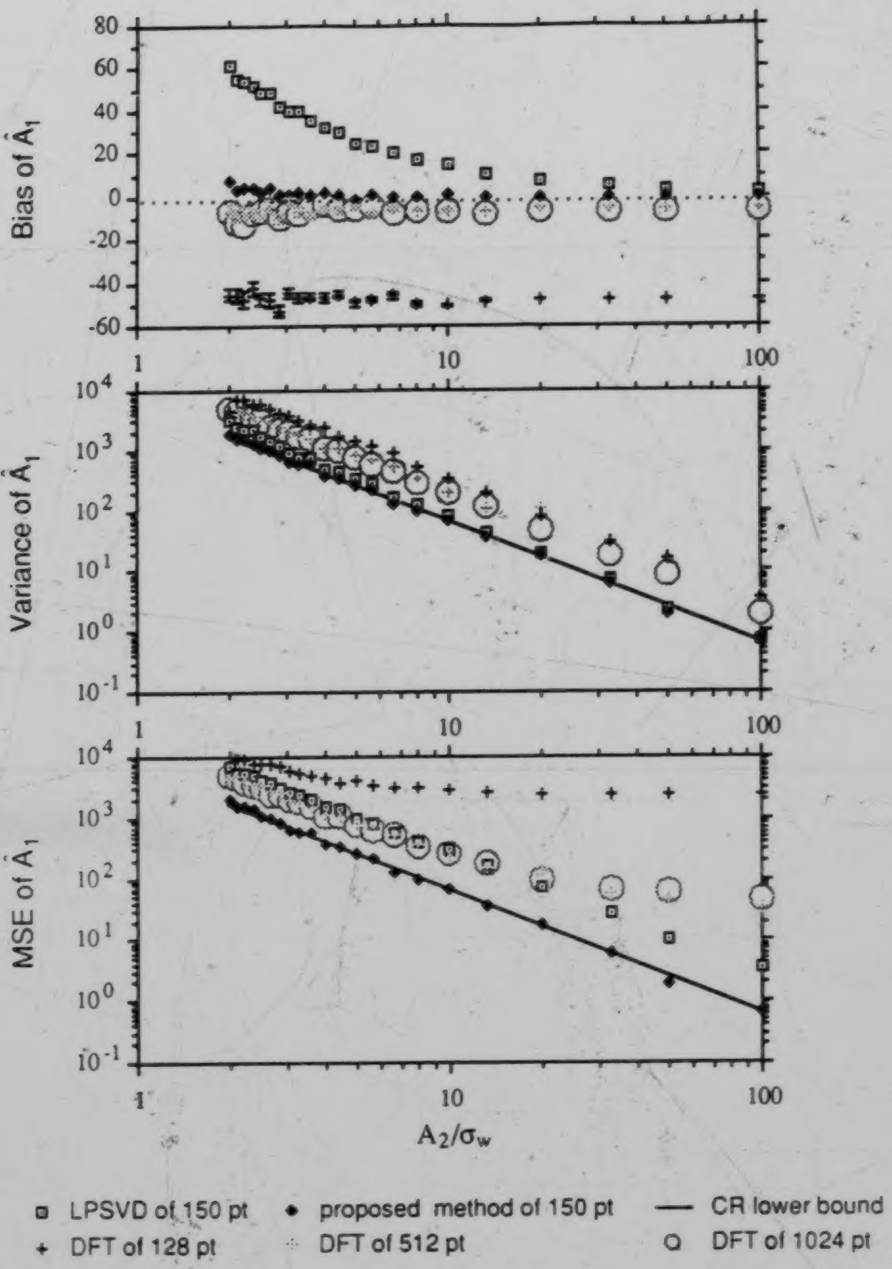


Figure 5.4.1.5 Time domain amplitude of the 1-st peak versus A_2/σ_w, where $A_2 = 250$. For other information, see Figure 5.4.1.4.

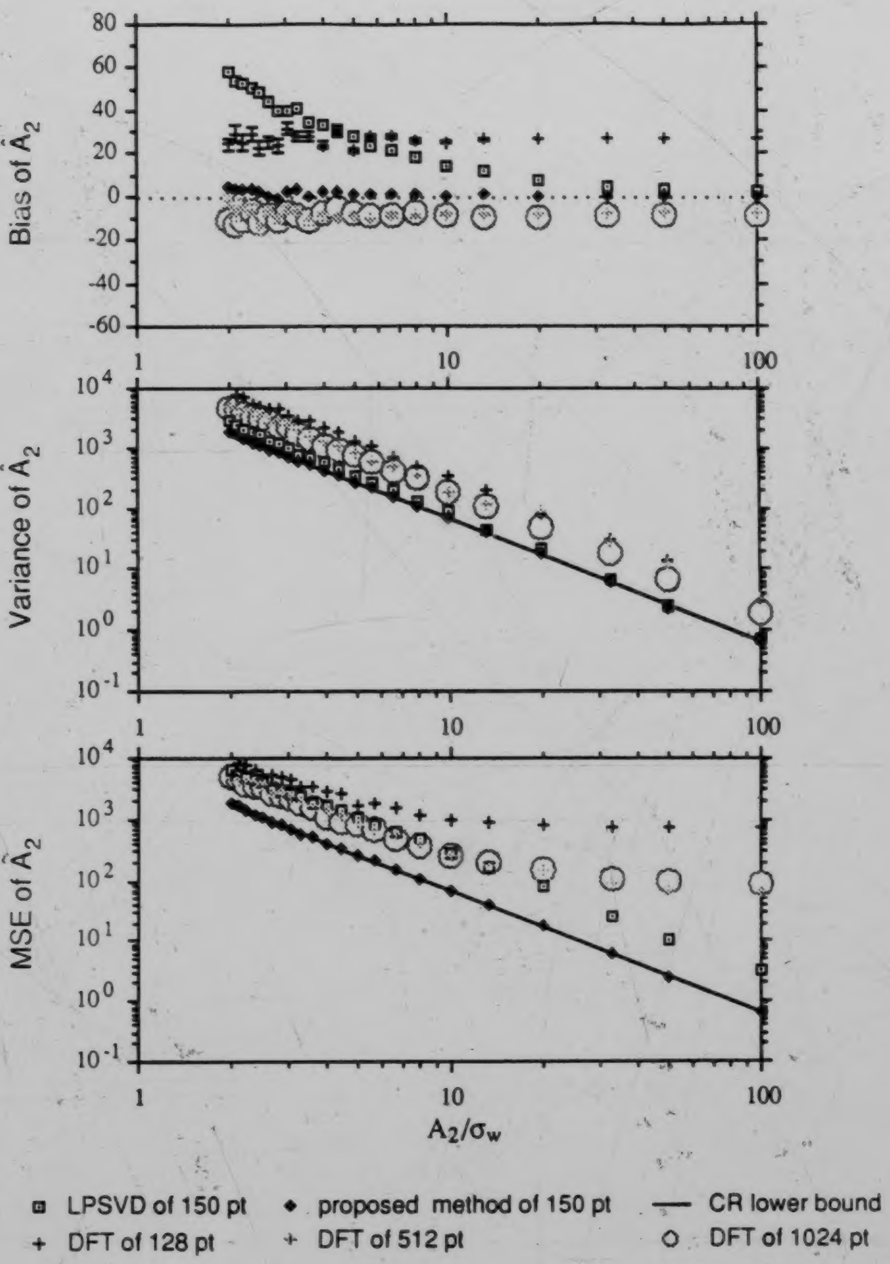


Figure 5.4.1.6 Time domain amplitude of the 2-nd peak versus A_2/σ_w, where A_2 = 250. For other information see Figure 5.4.1.4.

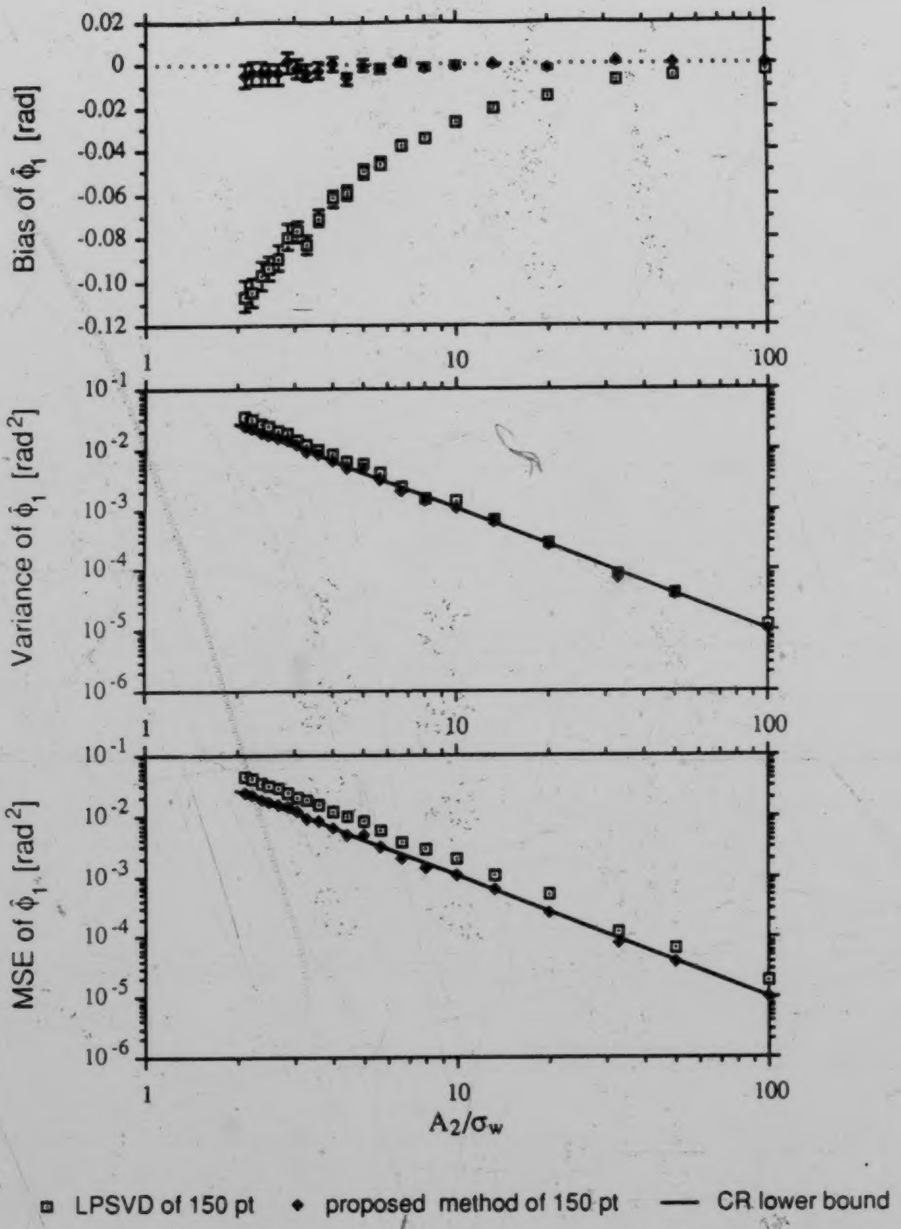


***Figure 5.4.1.7** Phase of the 1-st peak versus A_2/σ_w, where $A_2 = 250$. For other information, see Figure 5.4.1.4.*

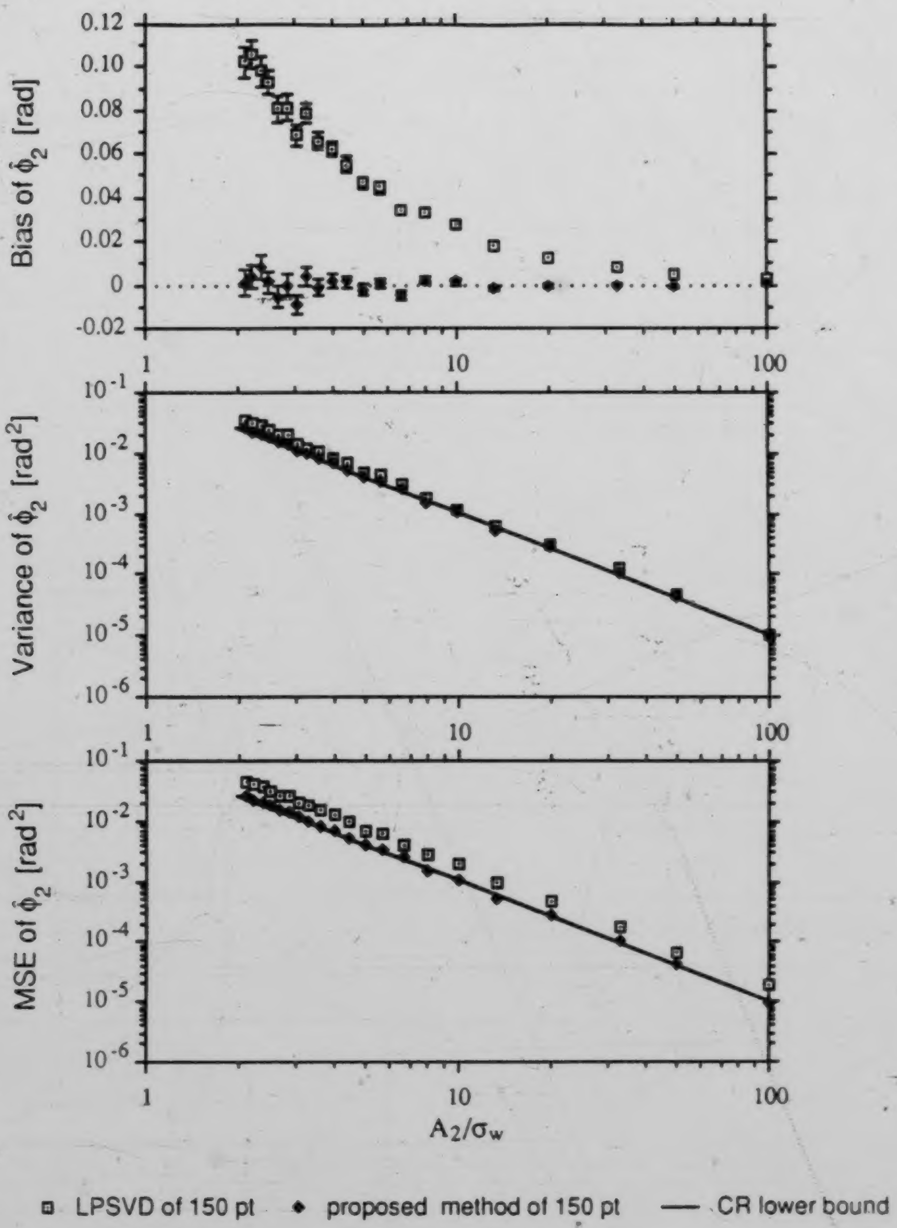


Figure 5.4.1.8 *Phase of the 2-nd peak versus A_2/σ_w, where $A_2 = 250$. For other information, see Figure 5.4.1.4.*

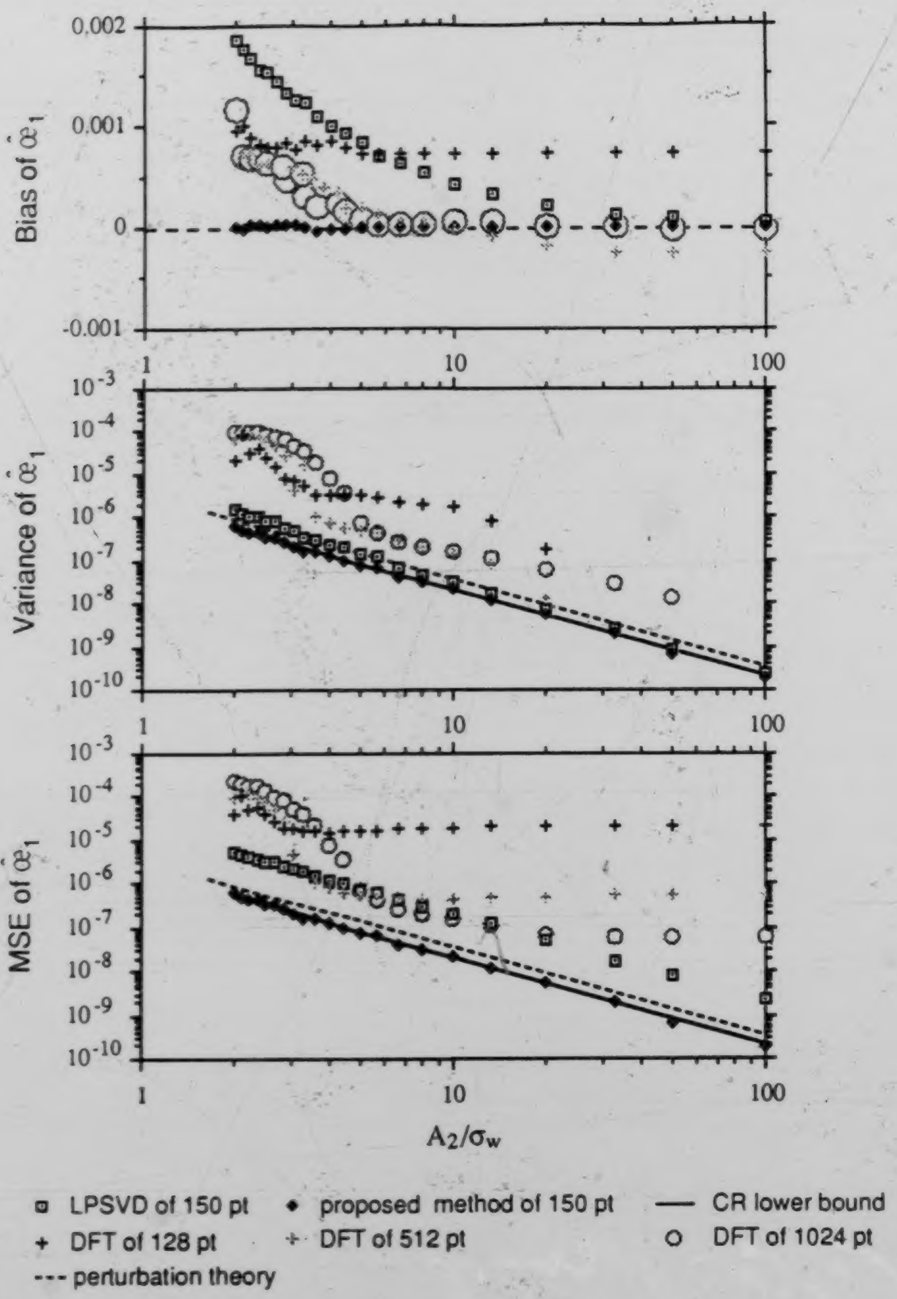


Figure 5.4.1.9 Normalized damping rate of the 1-st peak versus A_2/σ_w, where $A_2 = 250$. For other information, see Figure 5.4.1.4.

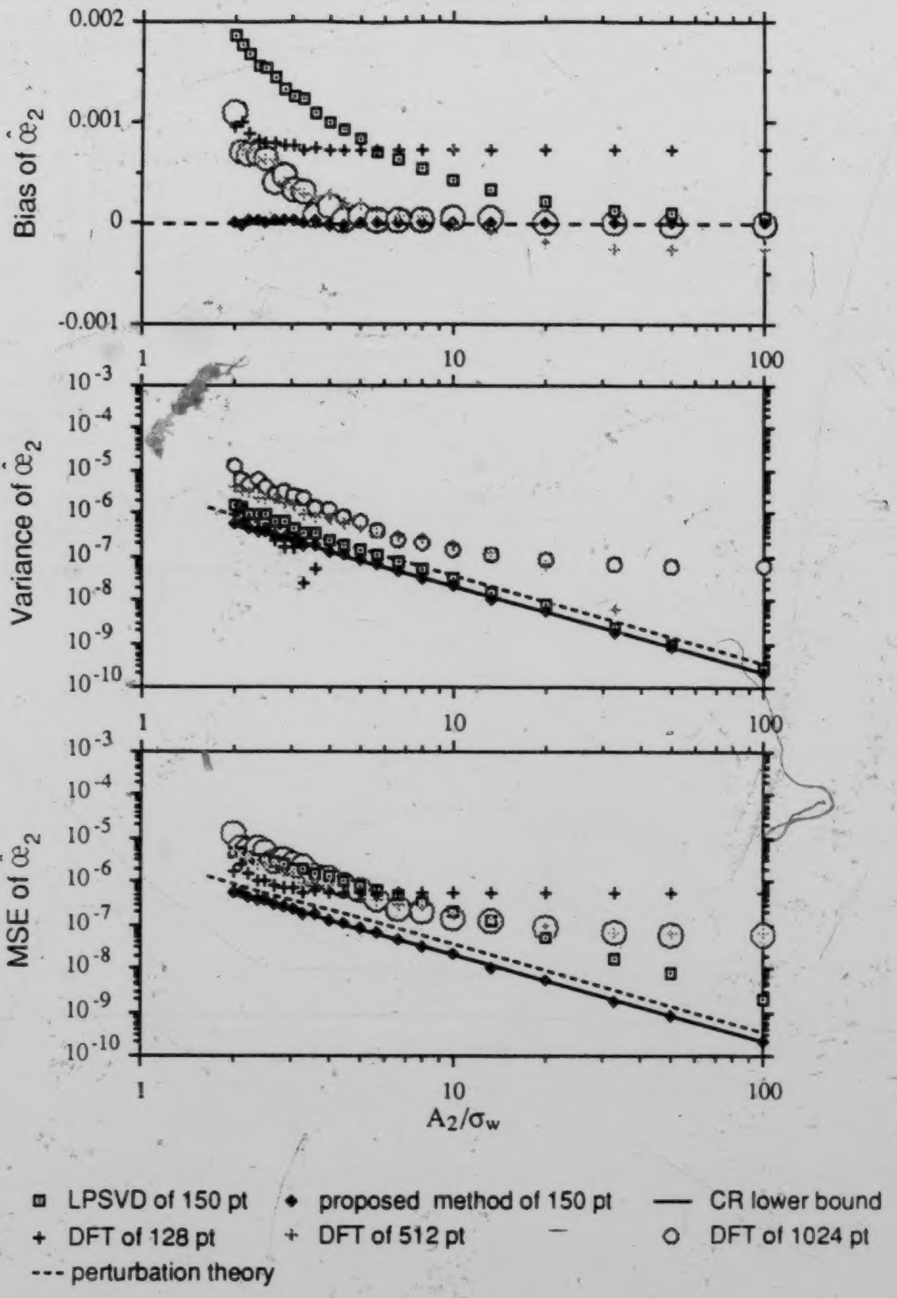


Figure 5.4.1.10 *Normalized damping rate of the 2-nd peak versus A_2/σ_w, where $A_2 = 250$. For other information, see Figure 5.4.1.4.*

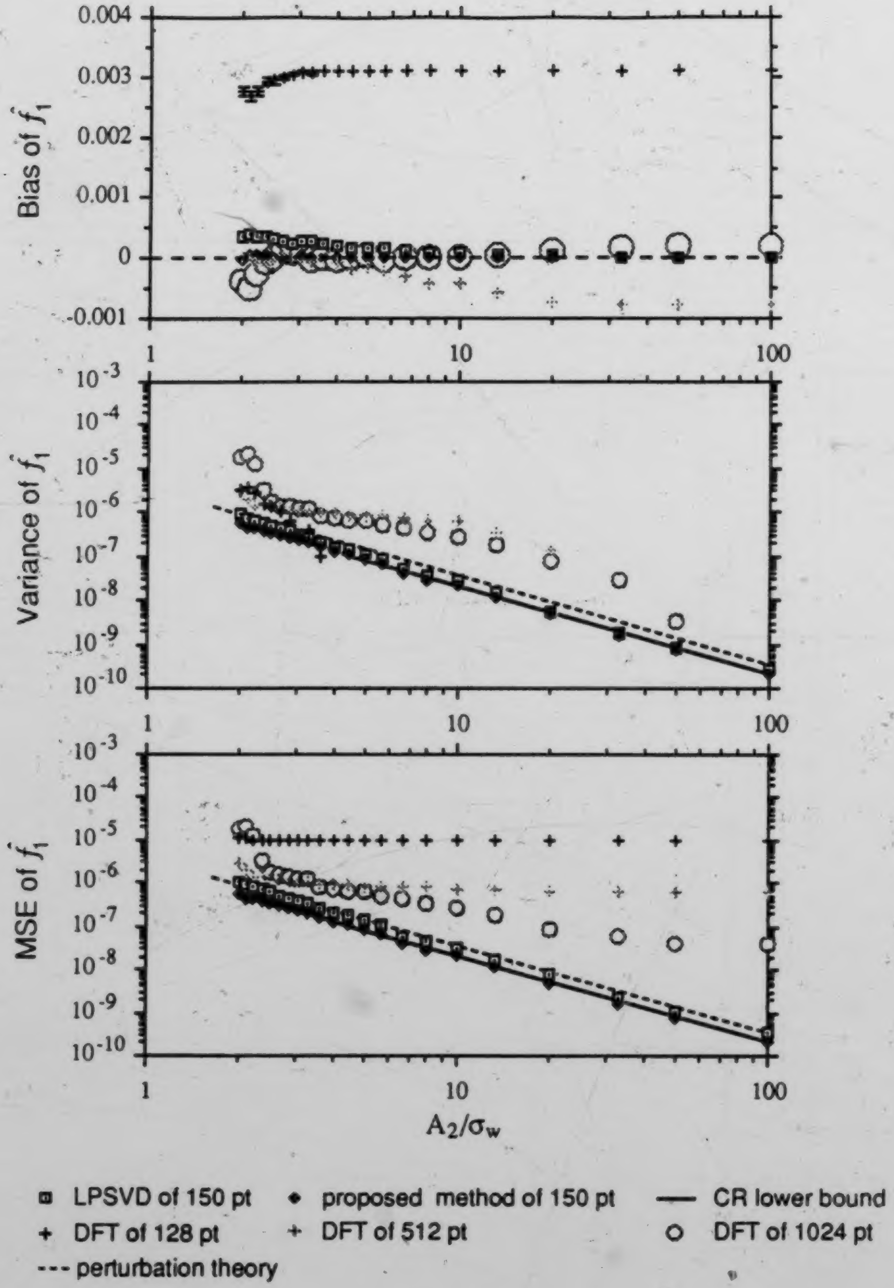


Figure 5.4.1.11 *Normalized frequency of the 1-st peak versus A_2/σ_w, where $A_2 = 250$. For other information, see Figure 5.4.1.4.*

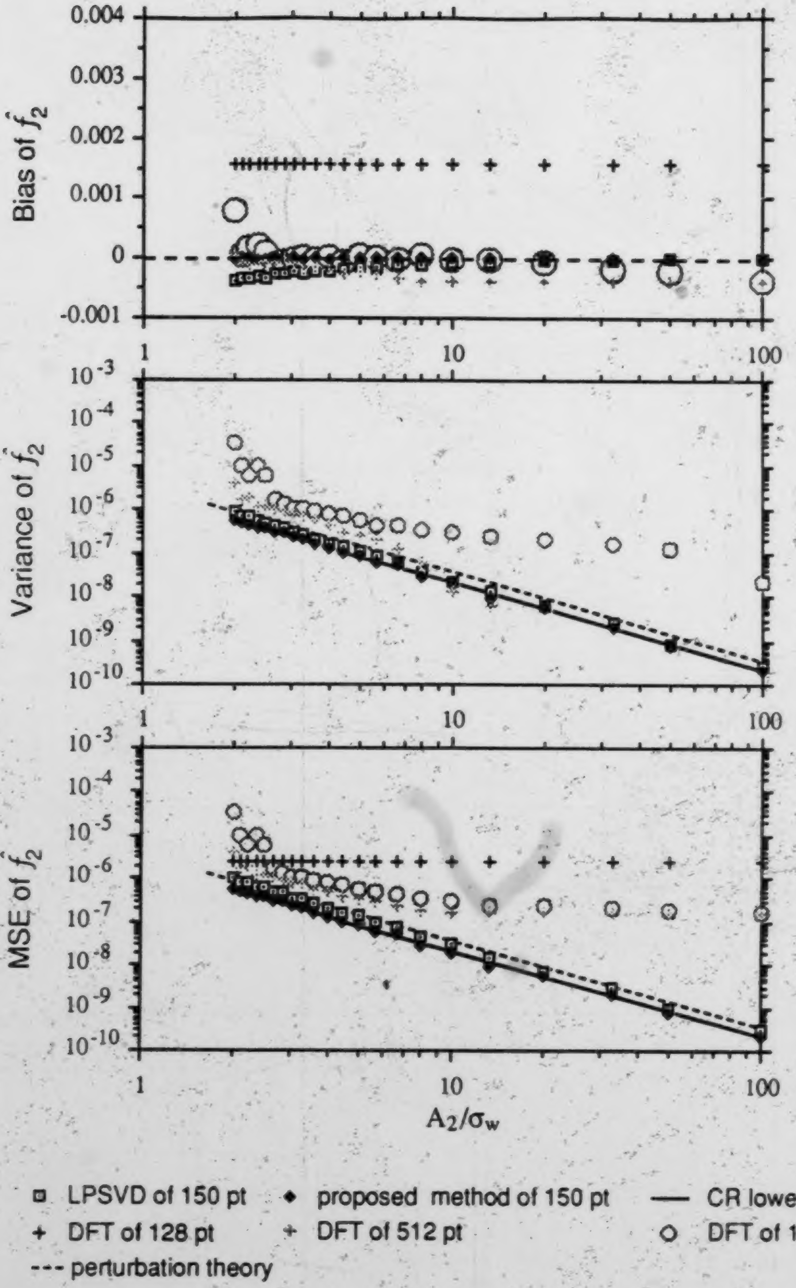


Figure 5.4.1.12 Normalized frequency of the 2-nd peak versus A_2/σ_w, where A_2 = 250. For other information, see Figure 5.4.1.4.

eventually become noise invariant at high SNRs. The LPSVD method of 150 points, the proposed method of 150 points, and the DFT method of 512 points and 1024 points, seem to give very similar results in the estimate of the ratio of peak areas. But, the DFT method with only 128 points behaved poorly.

5.5 Effect of peak separation (i.e., spectral resolution)

In this section, we will evaluate the effect of moving the location of one peak relative to a standard peak.

Example 5.5.1 :

Here we have two signals. The time domain amplitudes A_i, phase ϕ_i, normalized damping rate $œ_i$ and normalized frequencies f_i of the signals are tabulated as follows:

i	A_i	ϕ_i	$œ_i$	f_i
1	250	0	0.003183	-0.3
2	200	0	0.003183	f_2

where f_2 varies from -0.2 to -0.296. The sampling interval is kept at 4×10^{-4} sec and the location of the second peak is movable from the normalized frequency -0.2 to -0.296. In other words, the normalized frequency separation between the two peaks is changing from 0.1 to 0.004. Again over six hundred independent sequences of complex white Gaussian noise, each having independent real and imaginary parts of variance $\sigma_w^2 = (15)^2 = 225$, were added to the noise-free signals. The number of points used by the LPSVD method and the proposed method is 150, and the tentative model order for the LPSVD method is 75.

Figure 5.5.1.1 shows some representative spectra at different normalized frequency separations transformed by DFT from 1024 data points. Figure 5.5.1.2 shows the statistical results of the magnitude of the first five singular values from the LPSVD method. Figure 5.5.1.3 shows the magnitude of the first three singular values versus normalized frequency separation Δf. Here, we see that moving a peak's location changes both signal singular values nonlinearly. The noise singular values stay constant over the range of frequency separation studied in this example.

Figures 5.5.1.4 to 5.5.1.12 show the statistical results on the performance measures of the three methods for calculating the spectral parameters. For 128 data points, the DFT method cannot resolve the two peaks separated by less than the normalized frequency 0.01, while for 512 and 1024 data points, it cannot resolve the peaks separated by less than the normalized frequency 0.006. Thus, some point estimates from the DFT method are omitted in Figures 5.5.1.4, 5.5.1.6, 5.5.1.10 and 5.5.1.12. Recall from Appendix 3 that if the sampling frequency SW is 1000 Hz and the spectrometer frequency is 100 MHz (typical settings for ^1H spectroscopy), then normalized frequency separations of 0.01 and 0.006 are equivalent to 0.1 ppm and 0.06 ppm, respectively. On the other hand, if the sampling frequency is 2500 Hz and the spectrometer frequency is 40.55 MHz (typical settings for ^{31}P spectroscopy), then normalized frequency separations of 0.01 and 0.006 are equivalent to 0.62 ppm and 0.37 ppm, respectively.

As shown in Figures 5.5.1.4 to 5.5.1.12, both the LPSVD method and the proposed method become unstable (i.e., some estimates have large increases in biases and variances) when Δf is less than 0.004 (which is equivalent to 0.04 ppm and 0.25 ppm, respectively, for the above settings in ^1H and ^{31}P

normalized frequency separation, Δf	^{1}H @2T (SF=100MHz, SW=1000Hz) [ppm]	^{31}P @2T (SF=40.55MHz, SW=2500Hz) [ppm]
0.005	0.05	0.31
0.010	0.1	0.62
0.025	0.25	1.54
0.050	0.5	3.08

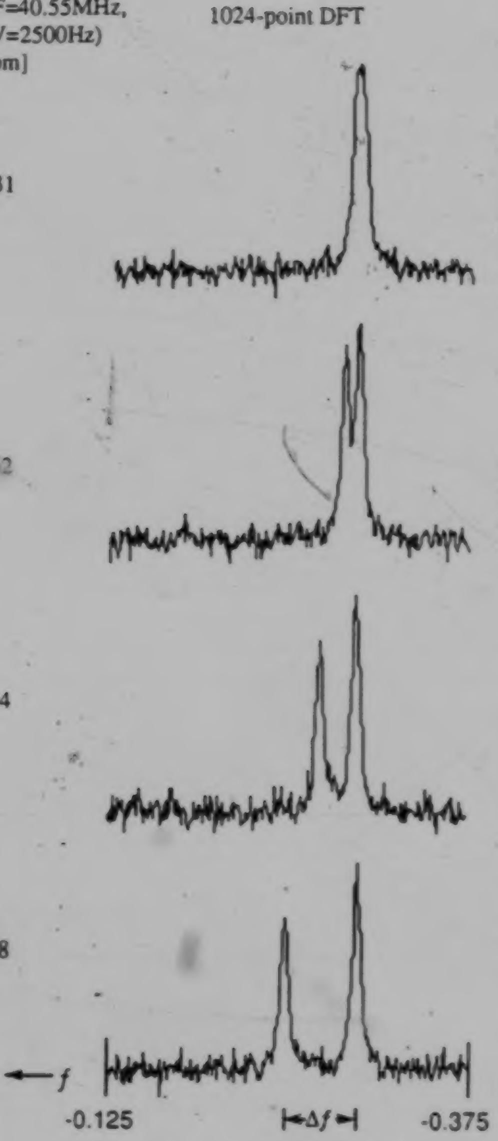


***Figure 5.5.1.1** Some representative 1024 data points DFT spectra at different normalized frequency separations, transformed from two sinusoids of time domain amplitudes 250 and 200, equal normalized damping rate (0.00318). The noise standard deviation in the time domain, as defined in Equation (1.1.1), is 15.*

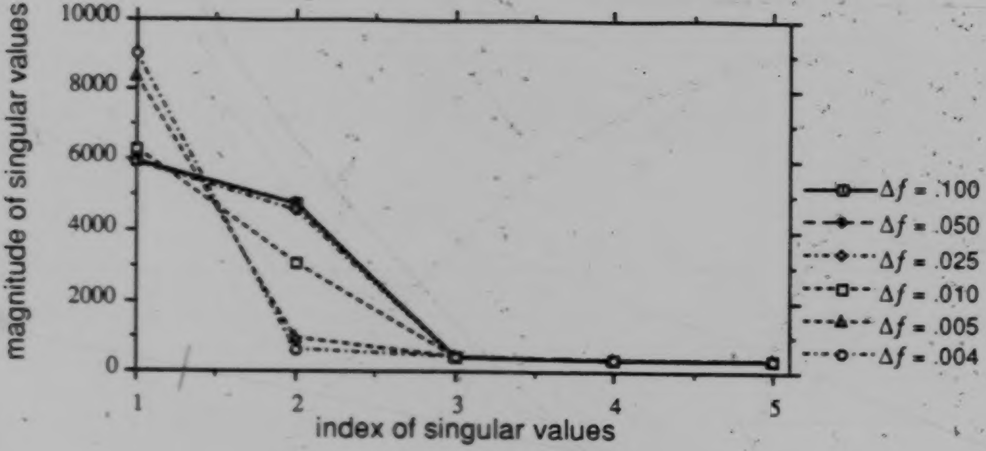


Figure 5.5.1.2 Statistical result of the first five singular values of the LPSVD method. Δf is the normalized frequency separation between the two peaks. Number of data points used is 150. The tentative model order L is 75. For other information, see Figure 5.5.1.1.

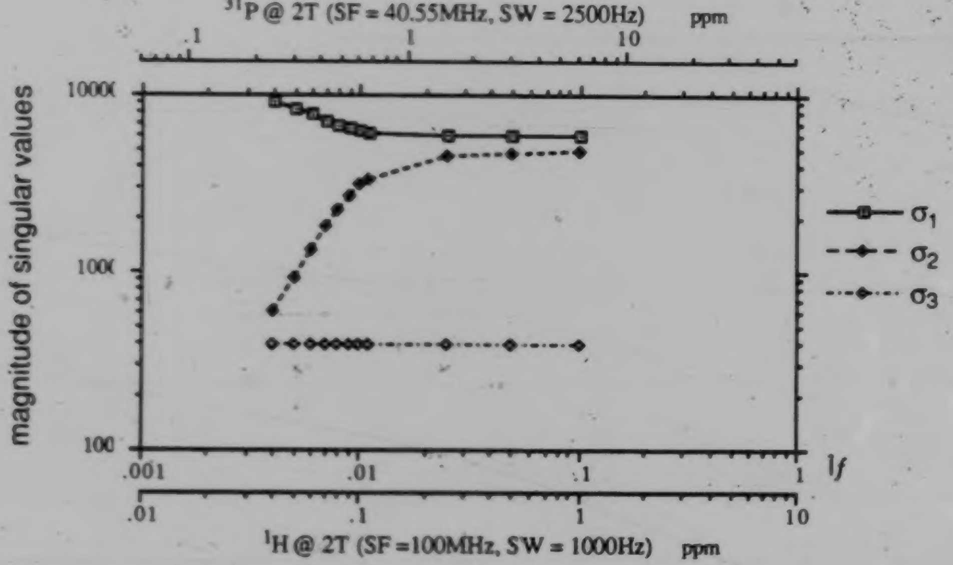


Figure 5.5.1.3 Magnitude of the first three singular values of the LPSVD method versus normalized frequency separation between two peaks Δf. For other information, see Figure 5.5.1.2.

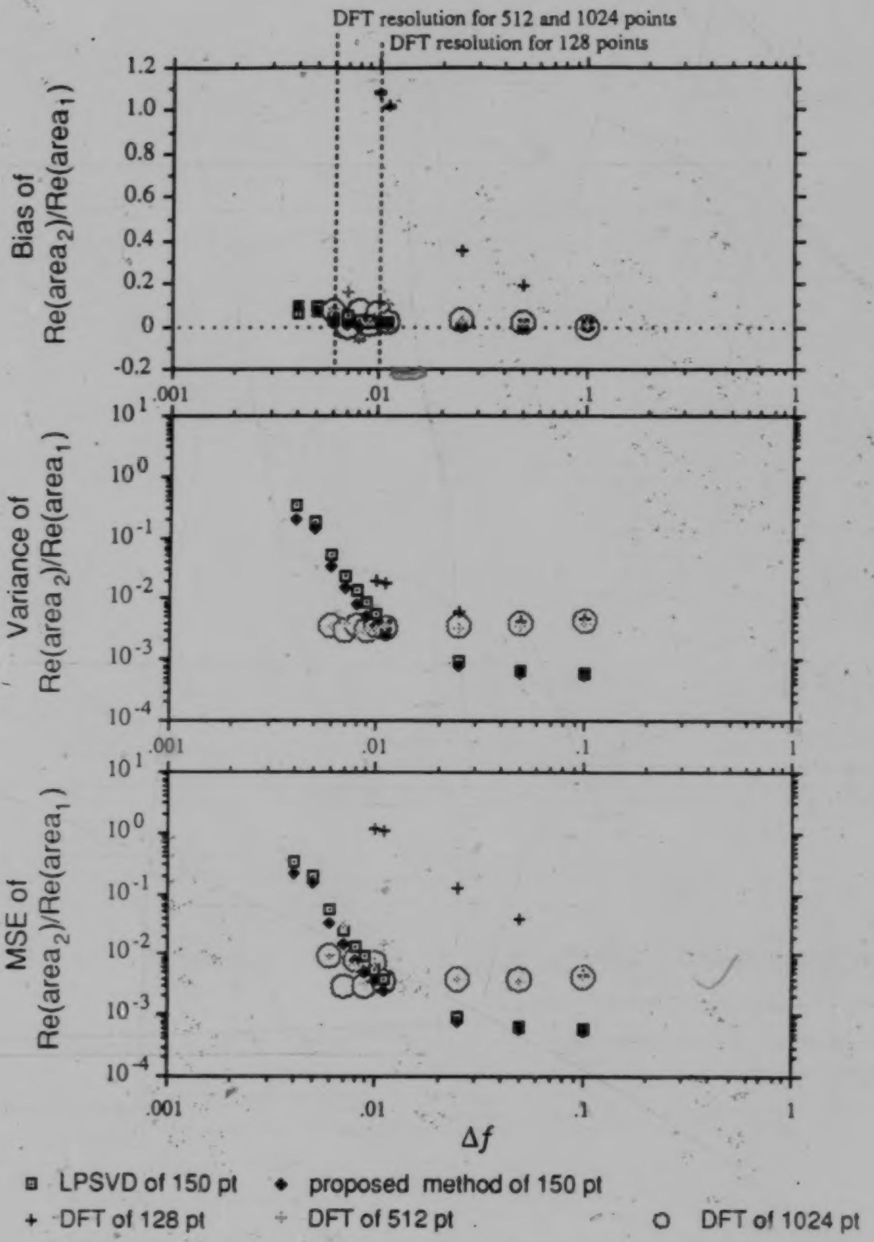


Figure 5.5.1.4 The area ratio of the two peaks from the real spectrum versus Δf, where Δf is the normalized frequency separation between the two peaks. $A_1 = 250$, $A_2 = 200$; $\phi_1 = \phi_2 = 0$; $\oe_1 = \oe_2 = 0.00318$. The tentative model order for the LPSVD method is 75. All estimates are obtained from over 600-run simulations.

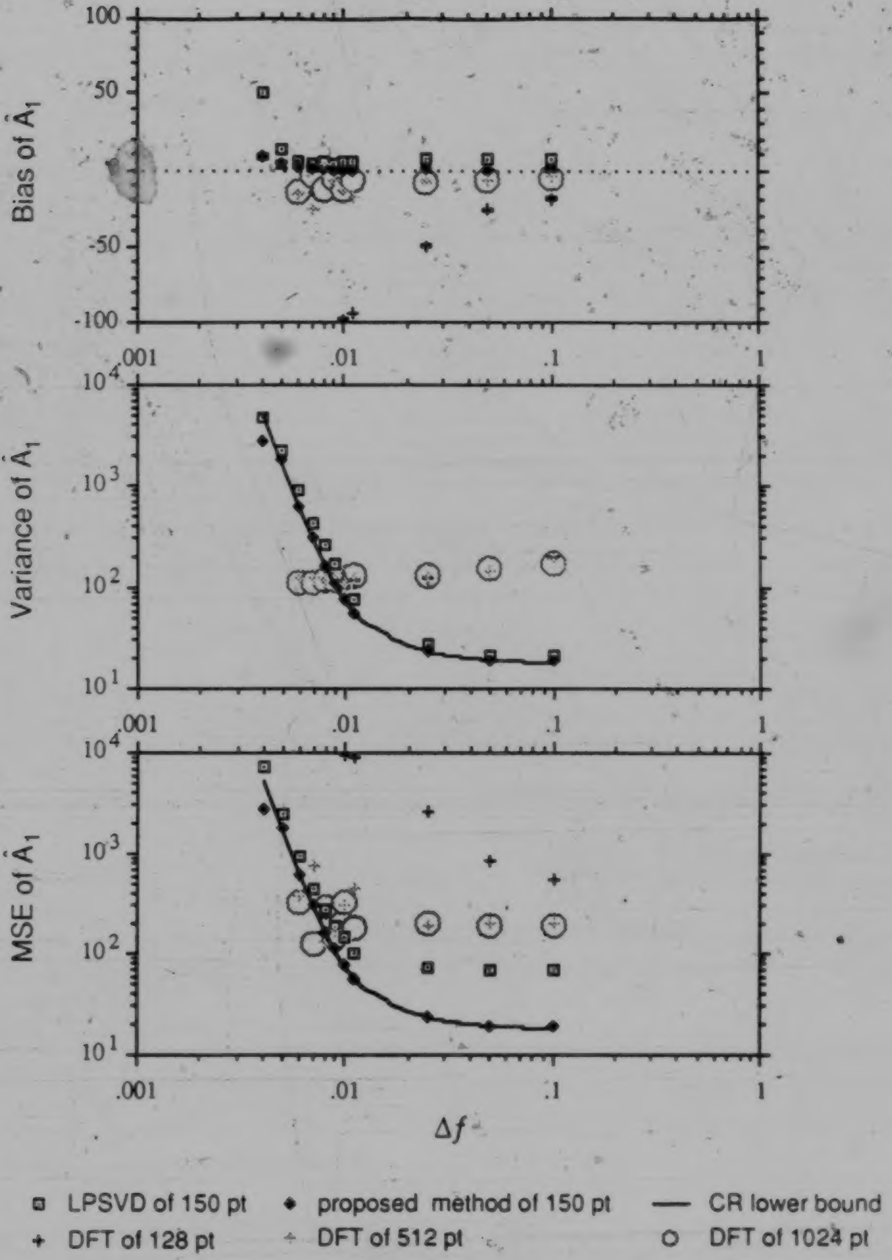


Figure 5.5.1.5 Time domain amplitude of the 1-st peak versus normalized frequency separation between the two peaks, Δf. For other information, see Figure 5.5.1.4.

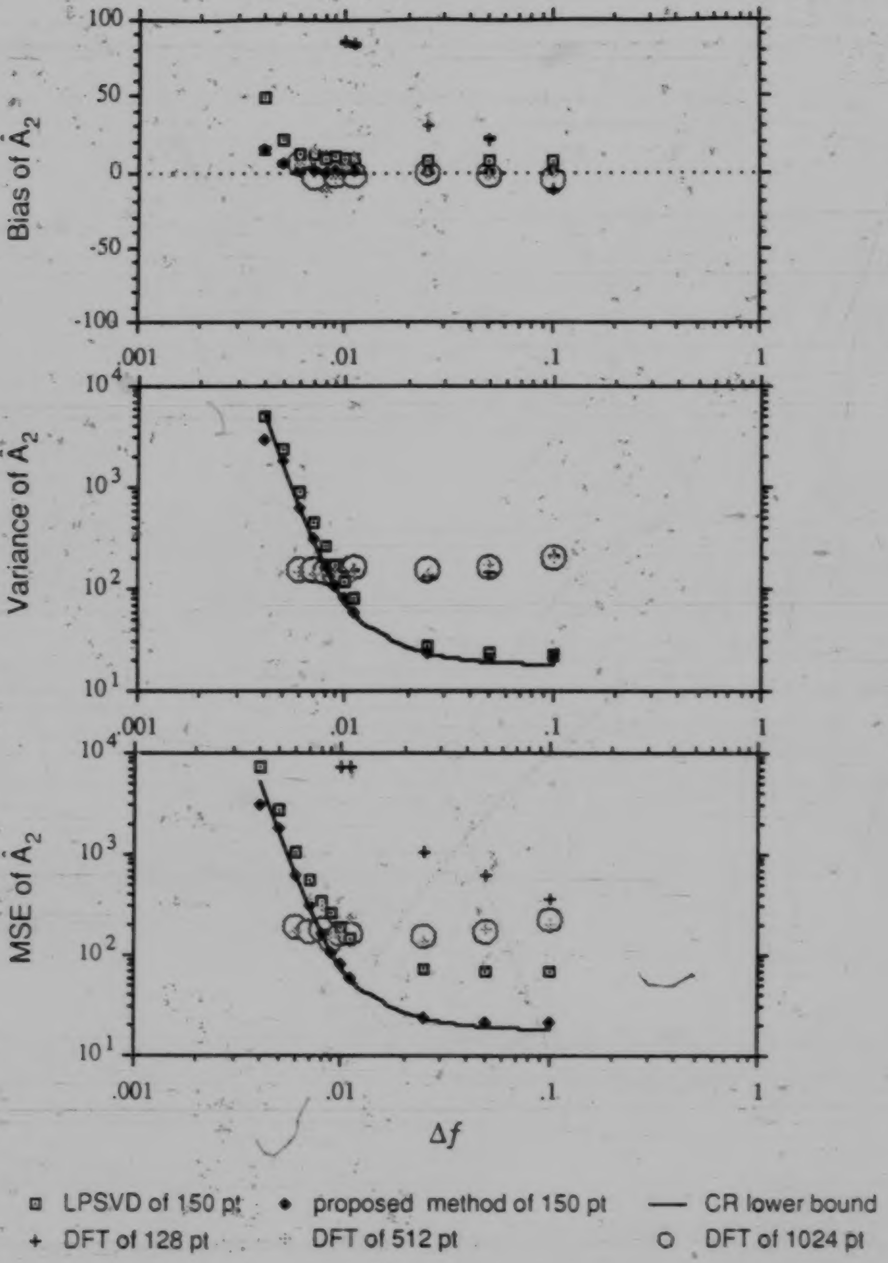


***Figure 5.5.1.6** Time domain amplitude of the 2-nd peak versus normalized frequency separation between the two peaks, Δf. For other information, see Figure 5.5.1.4.*

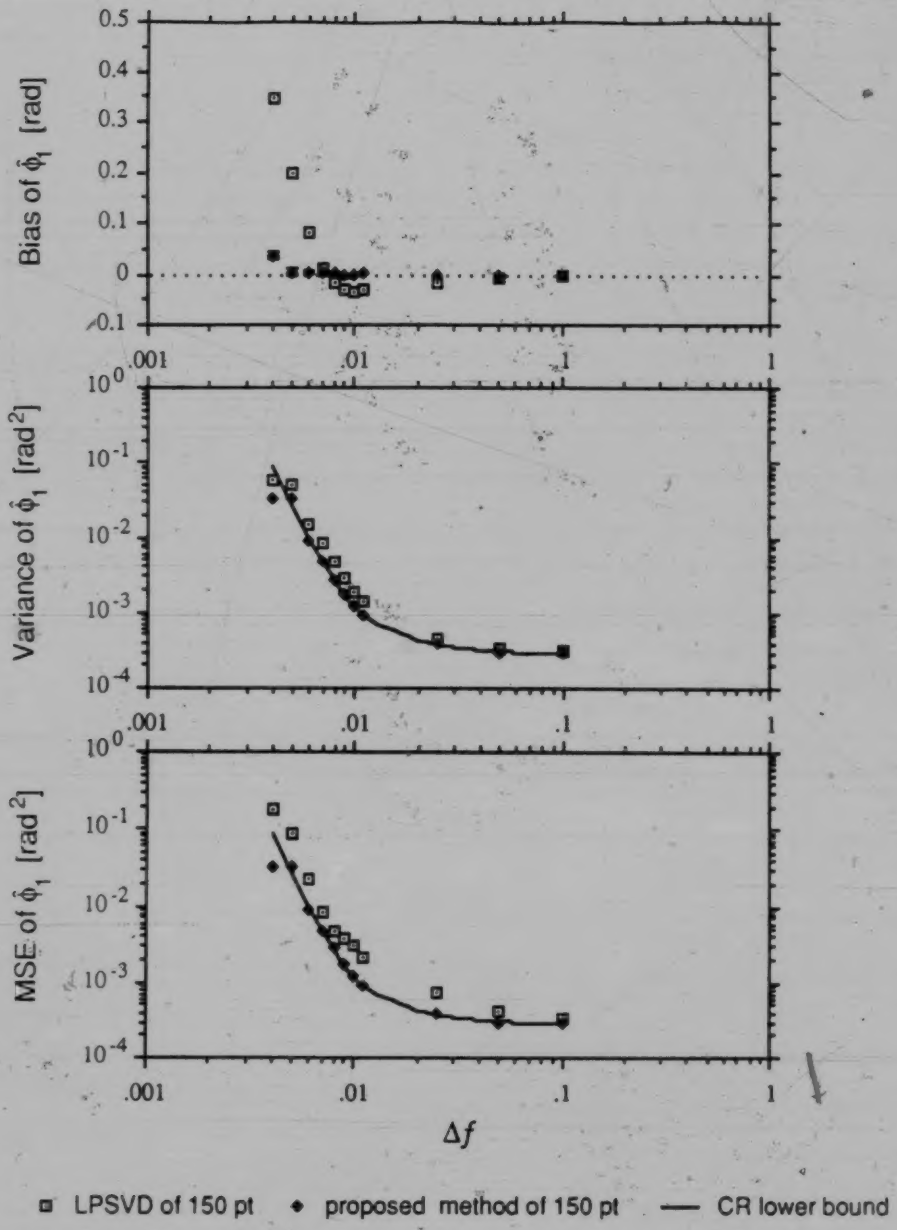
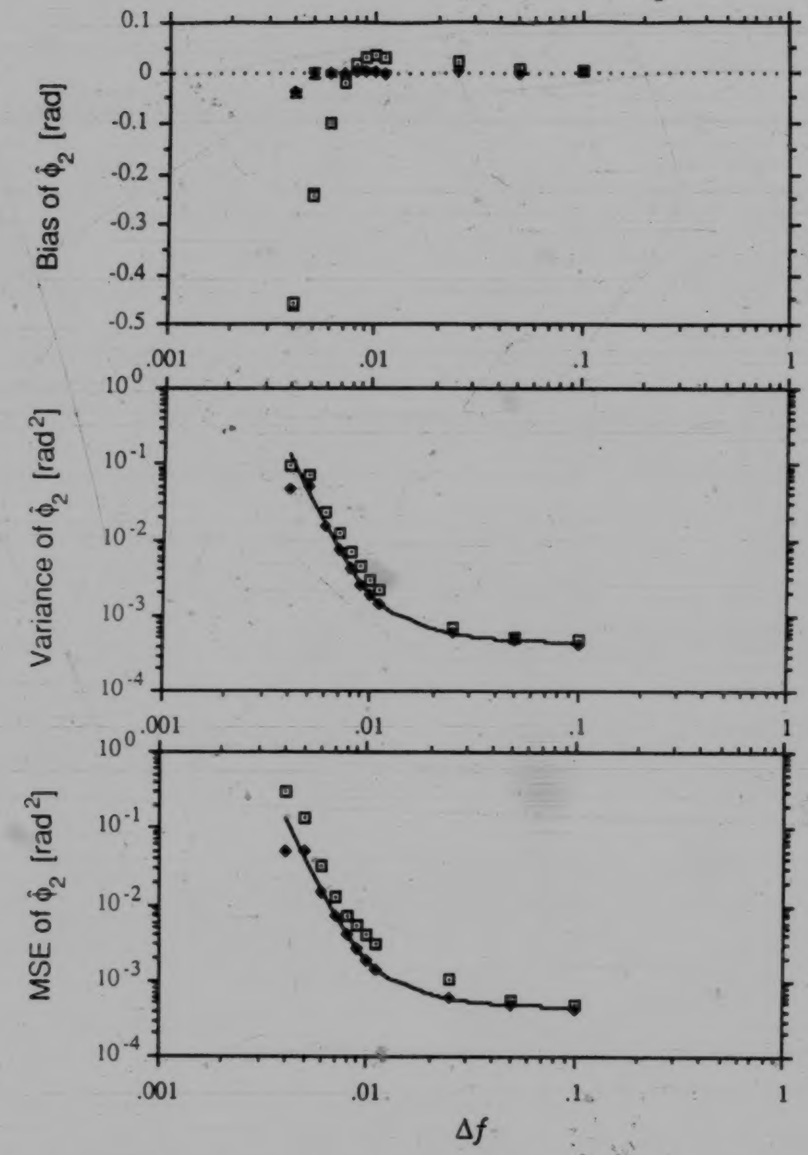


***Figure 5.5.1.7** Phase of the 1-st peak versus normalized frequency separation between the two peaks, Δf. For other information, see Figure 5.5.1.4.*



▣ LPSVD of 150 pt ◆ proposed method of 150 pt — CR lower bound

***Figure 5.5.1.8** Phase of the 2-nd peak versus normalized frequency separation between the two peaks, Δf. For other information, see Figure 5.5.1.4.*

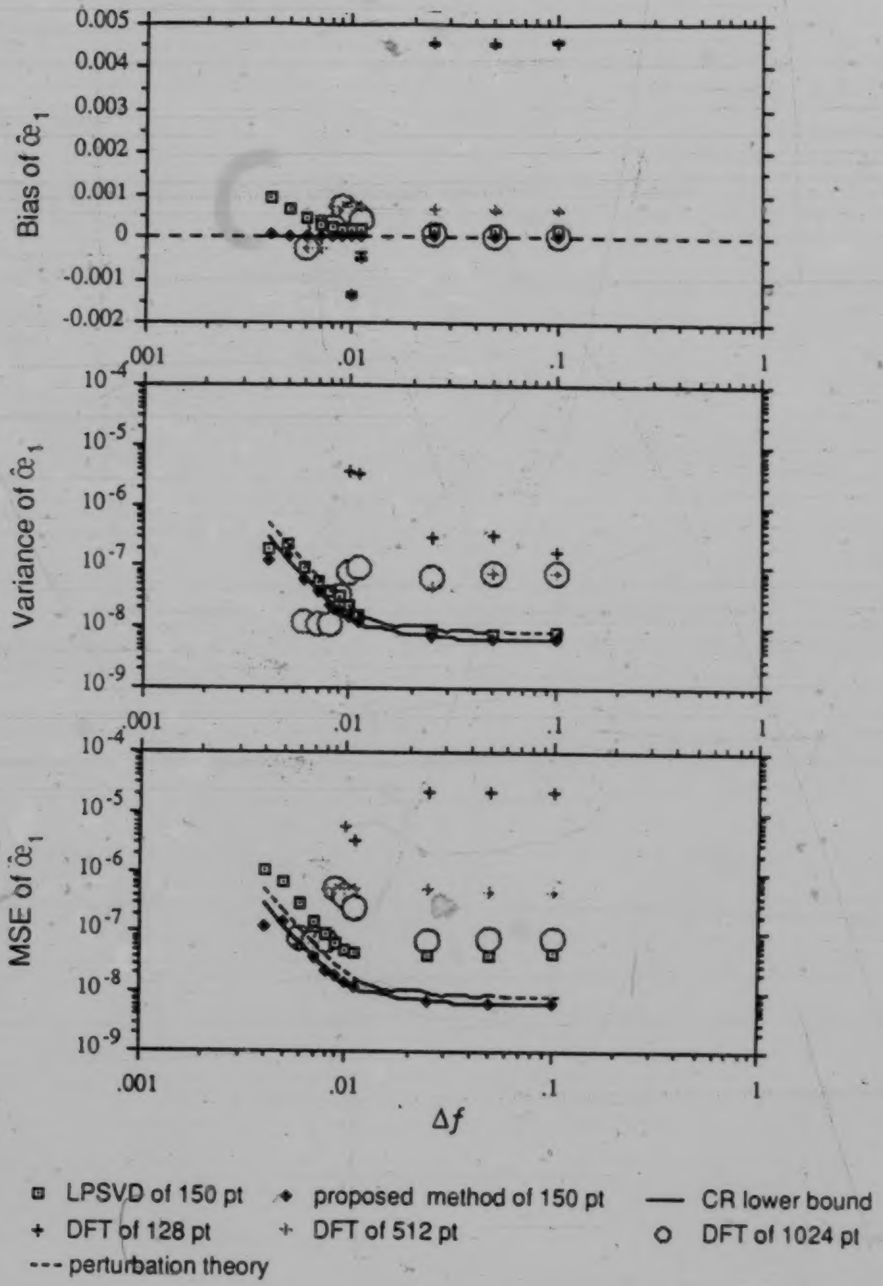


Figure 5.5.1.9 Normalized damping rate of the 1-st peak versus normalized frequency separation between the two peaks, Δf. For other information, see Figure 5.5.1.4.

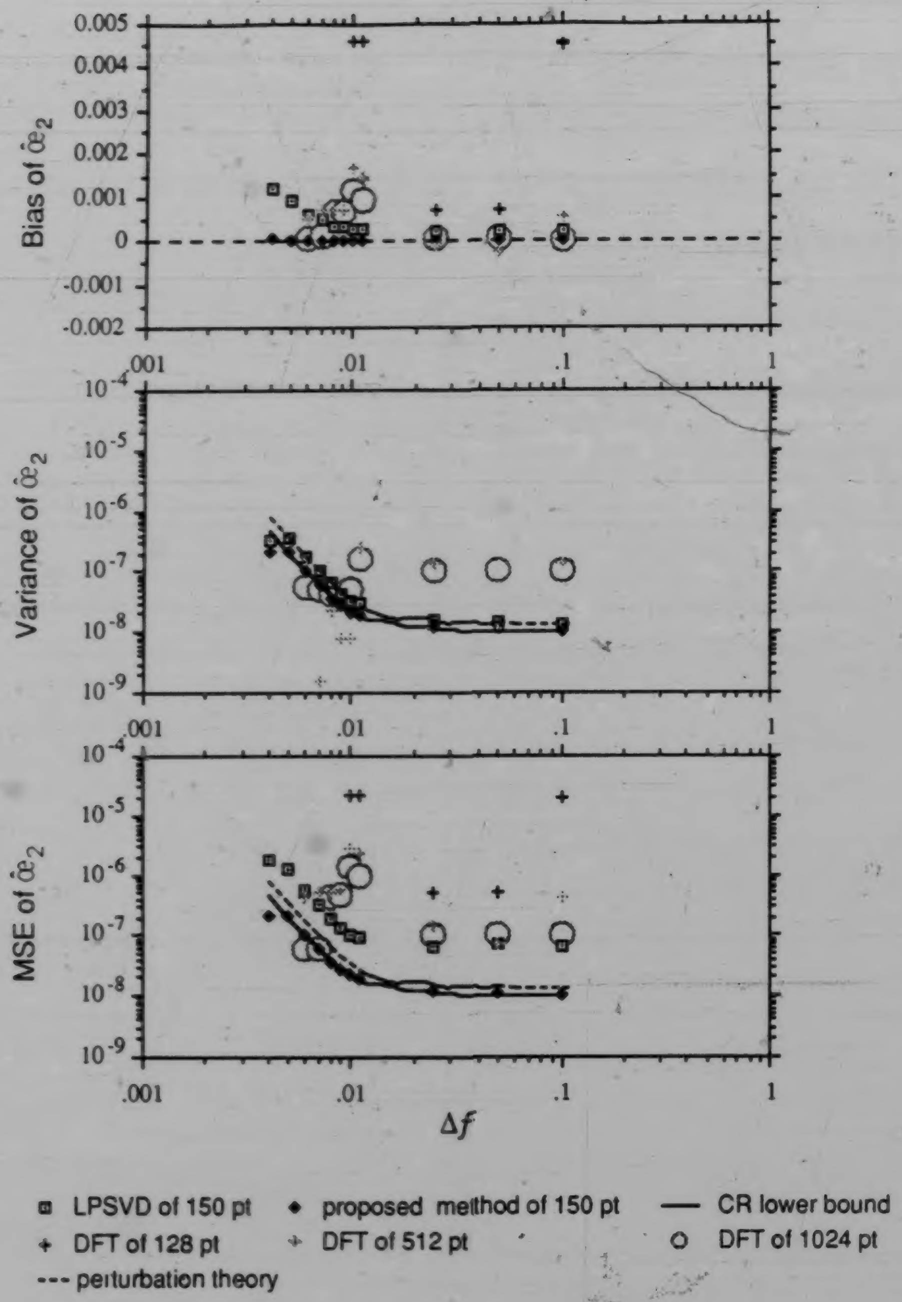


Figure 5.5.1.10 *Normalized damping rate of the 2-nd peak versus normalized frequency separation between the two peaks, Δf. For other information, see Figure 5.5.1.4.*

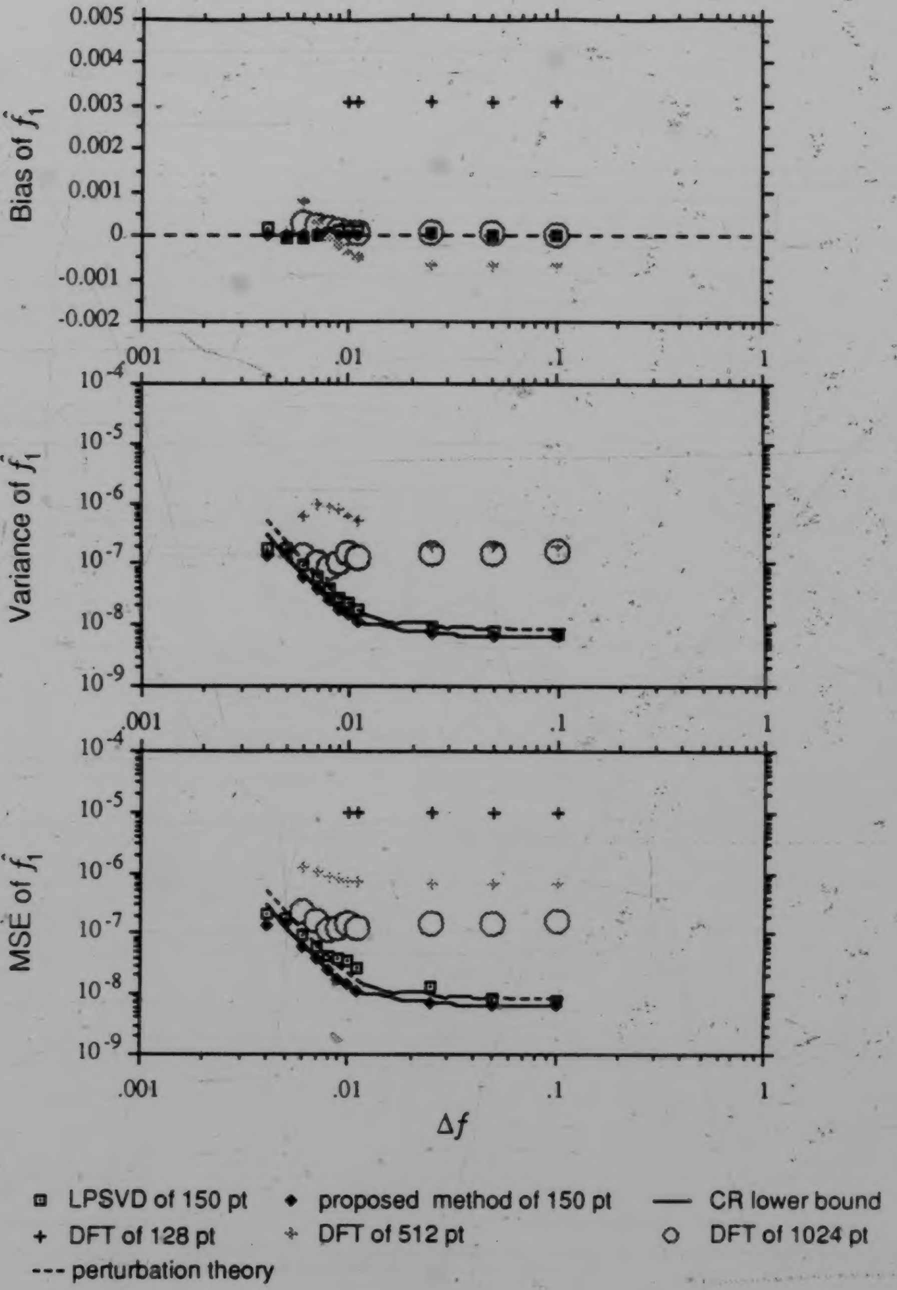


Figure 5.5.1.11 Normalized frequency of the 1-st peak versus normalized frequency separation between the two peaks, Δf. For other information, see Figure 5.5.1.4.

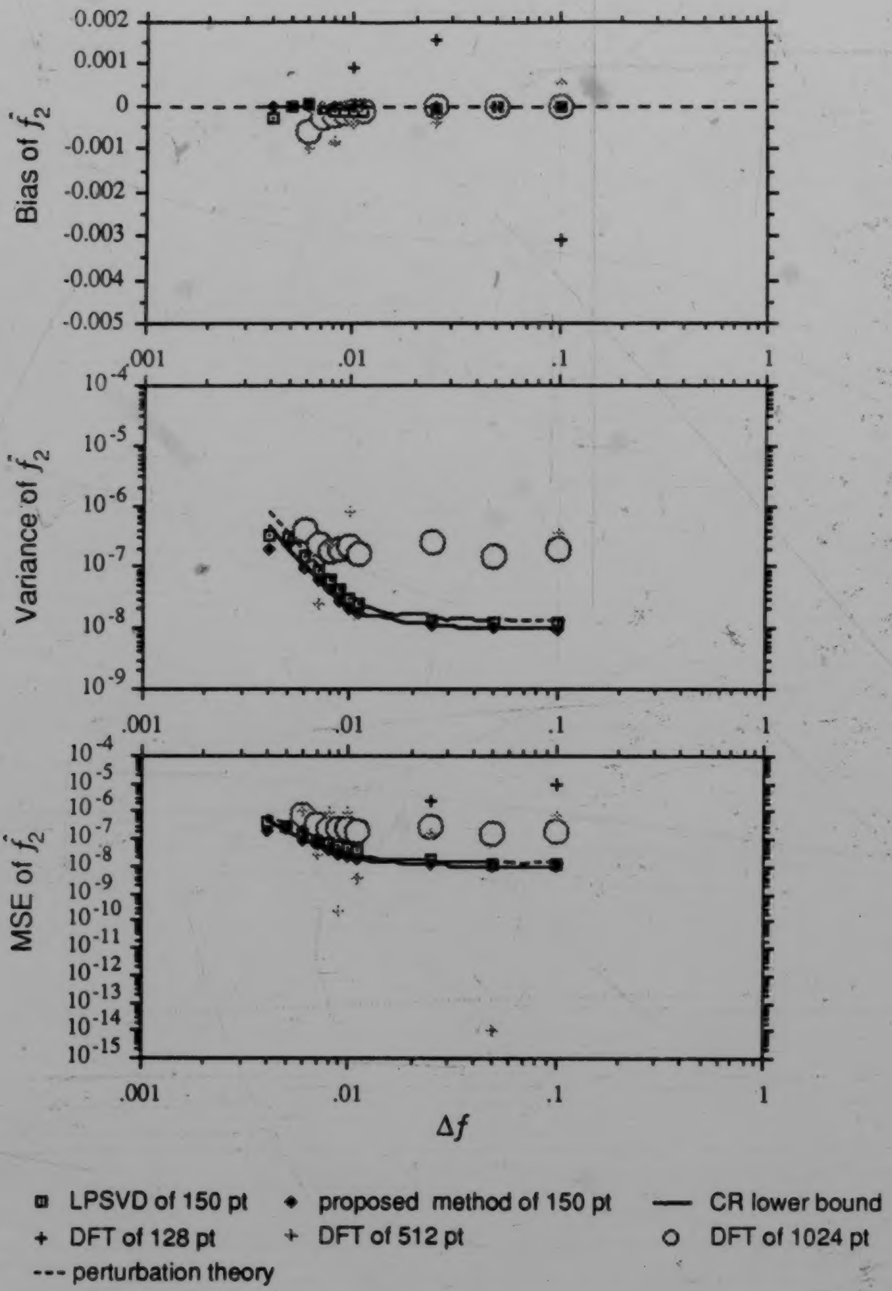


Figure 5.5.1.12 *Normalized frequency of the 2-nd peak versus normalized frequency separation between the two peaks, Δf. For other information, see Figure 5.5.1.4.*

spectroscopy). The results from the proposed method agree very well with the CR lower bounds. In addition, frequency and damping rate estimates from the LPSVD method are close to the results obtained from perturbation analysis in section 4.2.2. As expected, the LPSVD method and the proposed method produced very similar results for the estimates of the first peak and that of the second peak. The biases of the parameters from the LPSVD method deteriorate as the two peaks get closer to one another.

In this example, we see that in general both the LSPVD method and the proposed method with only 150 points out perform the DFT method with 512 points and 1024 points, and both former methods have higher resolving power. The DFT method of 1024 data points behaves similar to that of 512 data points.

5.6 Effect of changing a peak's damping rate (i.e., T_2 dependence)

In this section, we want to see how a peak's damping rate (or linewidth) affects the LPSVD method, the proposed method and the DFT method.

Example 5.6.1 :

To demonstrate further the resolving power of the LPSVD method and the proposed method, we will create two closely spaced exponentially decaying sinusoids and allow the damping rate of one to change. The time domain amplitudes A_i, phase ϕ_i, normalized damping rate α_i and normalized frequencies f_i of the two signals are as follows:

i	A_i	ϕ_i	œ_i	f_i
1	250	0	0.003183	-0.3
2	200	0	œ_2	-0.29

where œ_2 varies from 0.003183 to 0.016. The sampling interval Δt is still 4×10^{-4} sec. The number of points used by the LPSVD method and the proposed method is still 150, and the tentative model order for the LPSVD method is 75. The input complex Gaussian noise standard deviation is 15.

Figure 5.6.1.1 shows some representative spectra at different normalized damping rates evaluated with the DFT for 512 and 1024 data points. The DFT of 128 data points cannot resolve the two peaks. But for 512 and 1024 points, the two peaks can barely be seen from the spectrum when œ_2 increases beyond about 0.0064.

Figure 5.6.1.2 shows the statistical result of the magnitude of the first five singular values. Figure 5.6.1.3 shows the magnitude of the first three singular values versus input œ_2. From this graph, we see that the noise singular values stay constant over the range of damping rates studied in this example. But, the second singular value σ_2 varies significantly with input œ_2.

Figures 5.6.1.4 to 5.6.1.12 show the statistical results of the estimates from the three methods. As mentioned above, the DFT method with 128 points cannot resolve the two peaks, so all figures related to the second peak do not have results from the 128-point DFT method. For 512 and 1024 points, the DFT method cannot resolve the peaks as œ_2 increases beyond 0.0064. But the LPSVD method

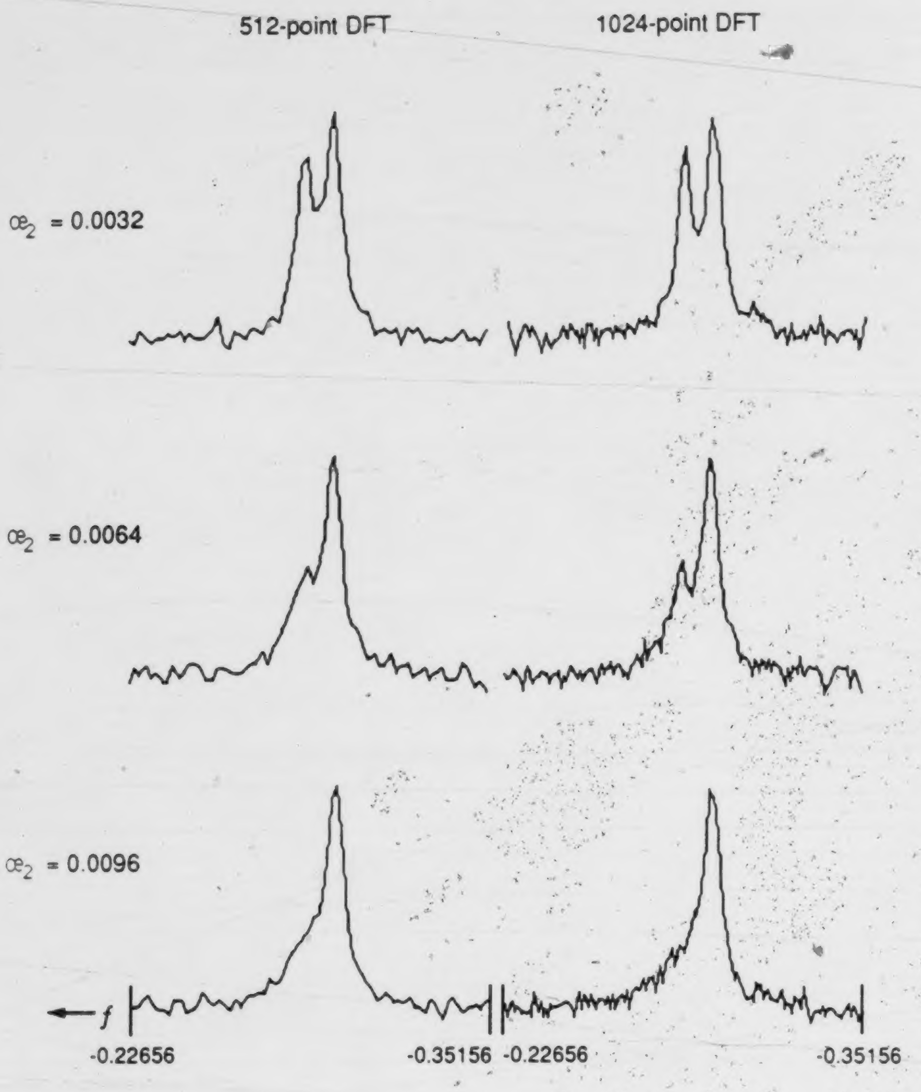


Figure 5.6.1.1 *Some representative DFT spectra acquired using different normalized damping rates for the 2-nd peak, transformed from 512 and 1024-point FID data which is composed of two decaying sinusoids plus Gaussian noise. The time domain parameters are : amplitudes $A_1 = 250$, $A_2 = 200$; normalized damping rate $\text{œ}_1 = 0.00318$; normalized frequencies $f_1 = -0.3$, $f_2 = -0.29$; noise standard deviation 15.*

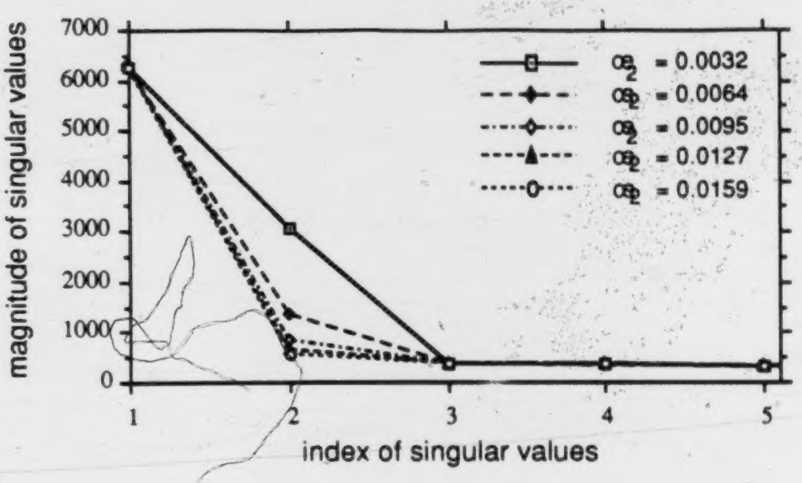


Figure 5.6.1.2 *Statistical result of the first five singular values of the LPSVD method. œ$_2$ is the normalized damping rate of the 2-nd peak. The number of data points used is 150. The tentative model order L is 75. For other information, see Figure 5.6.1.1.*

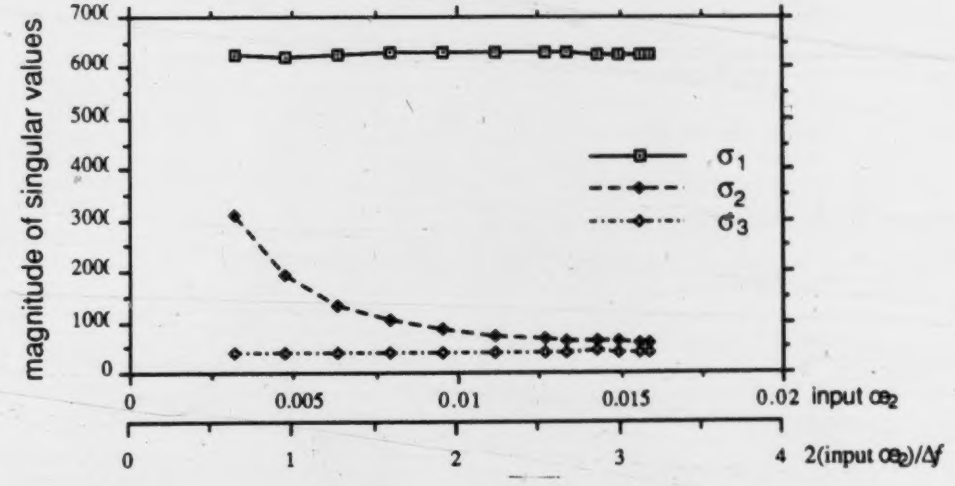


Figure 5.6.1.3 *Magnitude of the first three singular values of the LPSVD method versus normalized damping rate of the 2-nd peak, œ$_2$. For other information, see Figure 5.6.1.2.*

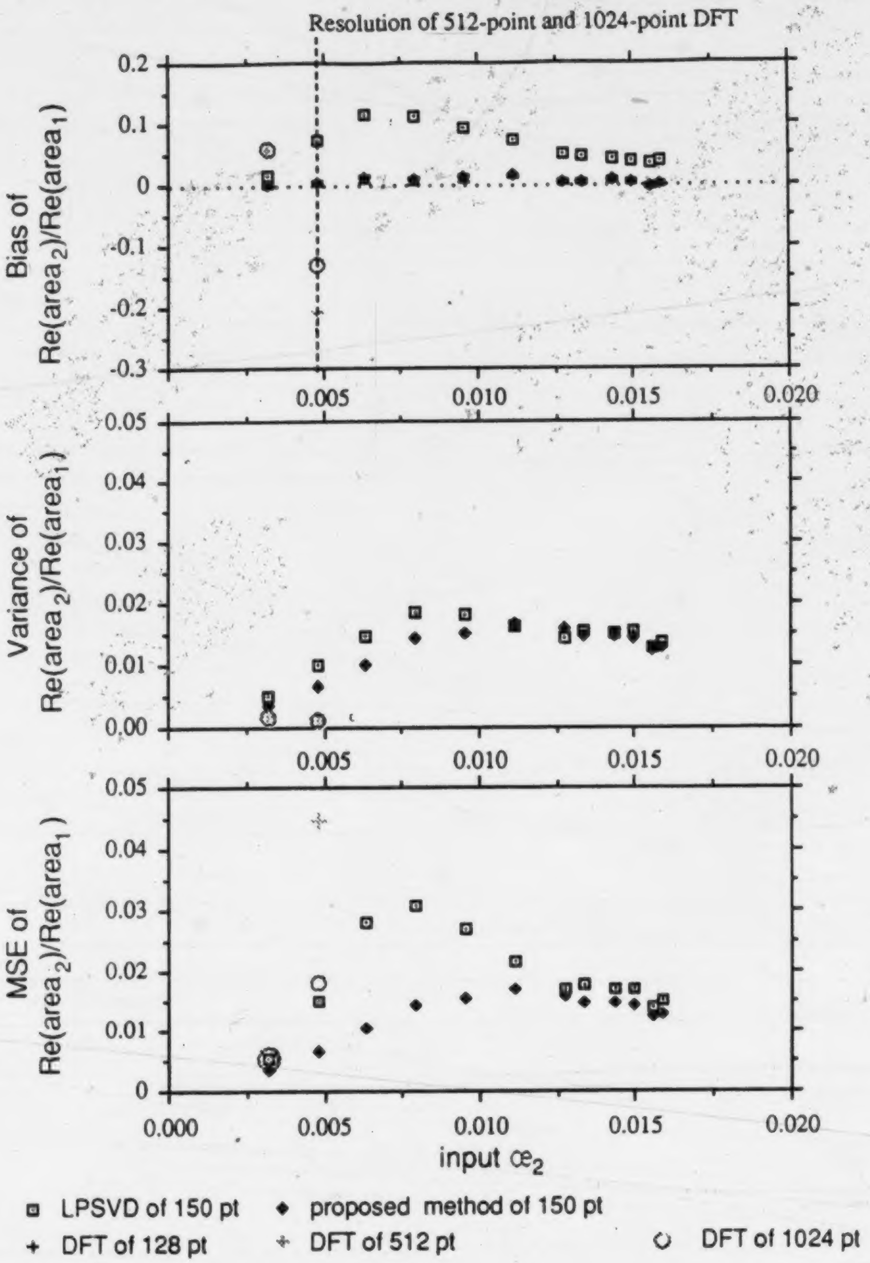


Figure 5.6.1.4 *The area ratio of the two peaks from the real spectrum versus input normalized damping rate of the 2-nd peak,* $\oe_2$. $A_1 = 250$, $A_2 = 200$; $\phi_1 = \phi_2 = 0$; $\oe_1 = 0.00318$; $f_1 = -0.3$, $f_2 = -0.29$; $\sigma_w = 15$. *Tentative model order for the LPSVD method is 75. All estimates are obtained from over 600-run simulations. For more detailed information, see Figure 5.6.1.2.*

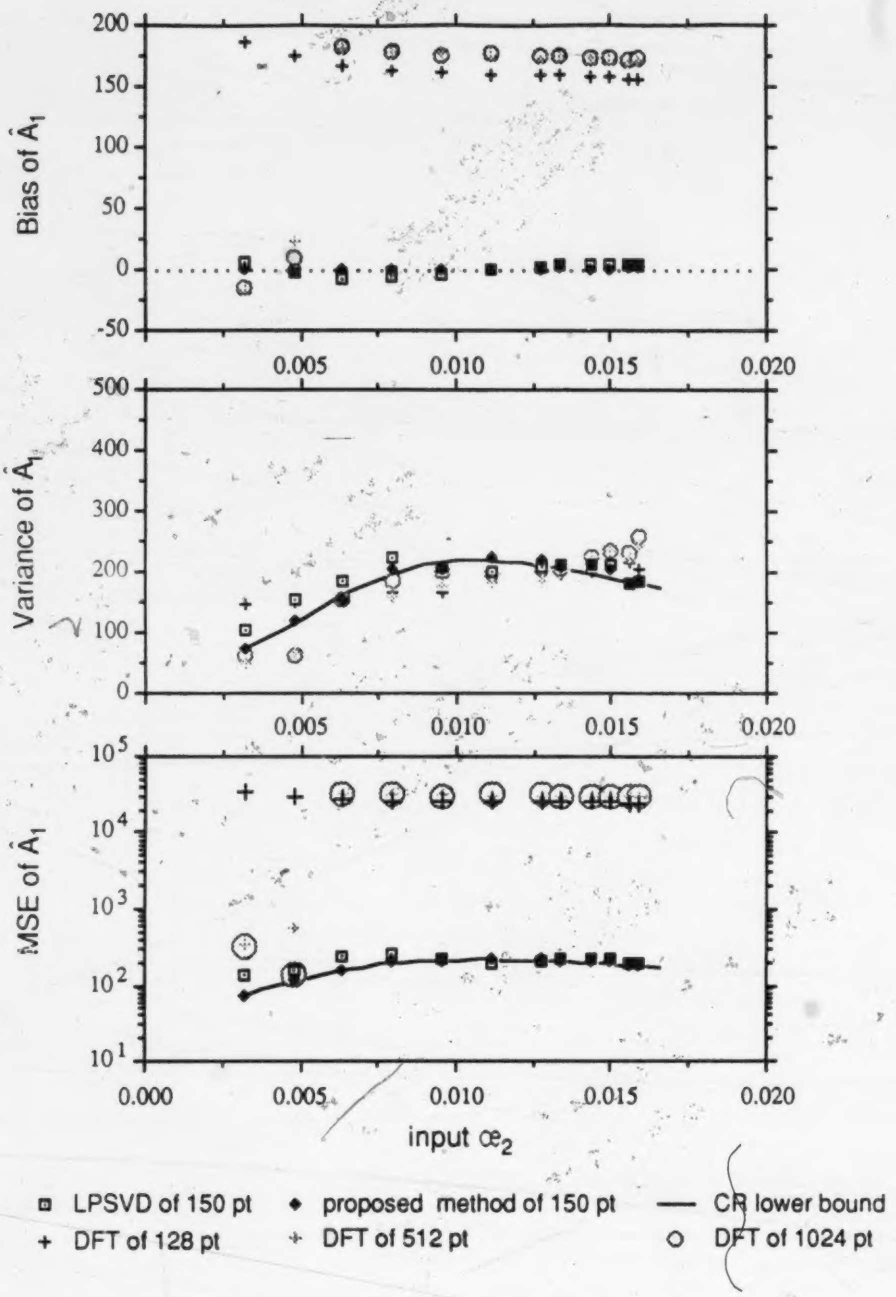


Figure 5.6.1.5 *Time domain amplitude of the 1-st peak versus input normalized damping rate of the 2-nd peak, α_2. For other information, see Figure 5.6.1.2.*

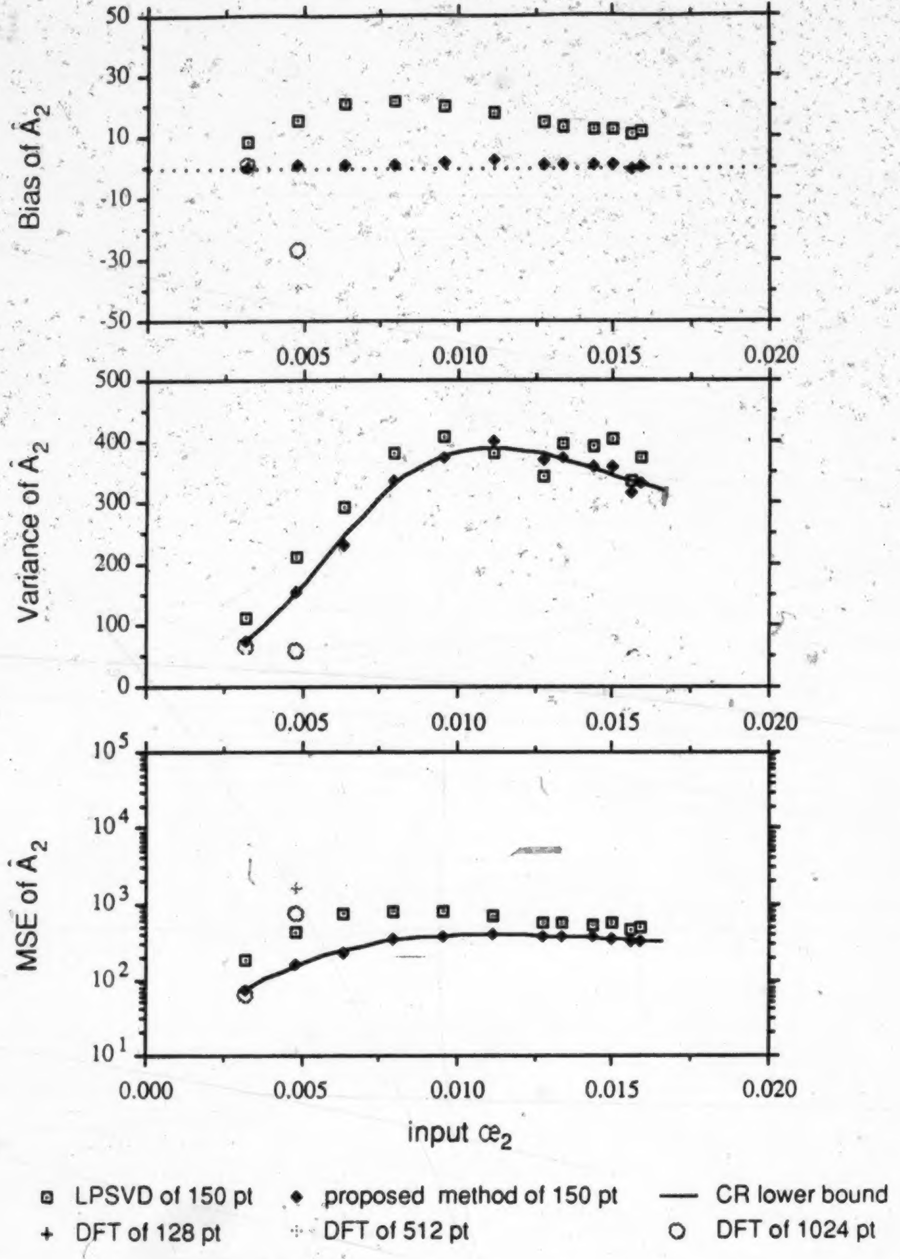


Figure 5.6.1.6 *Time domain amplitude of the 2-nd peak versus input normalized damping rate of the 2-nd peak,* $\oe_2$*. For other information, see Figure 5.6.1.2.*

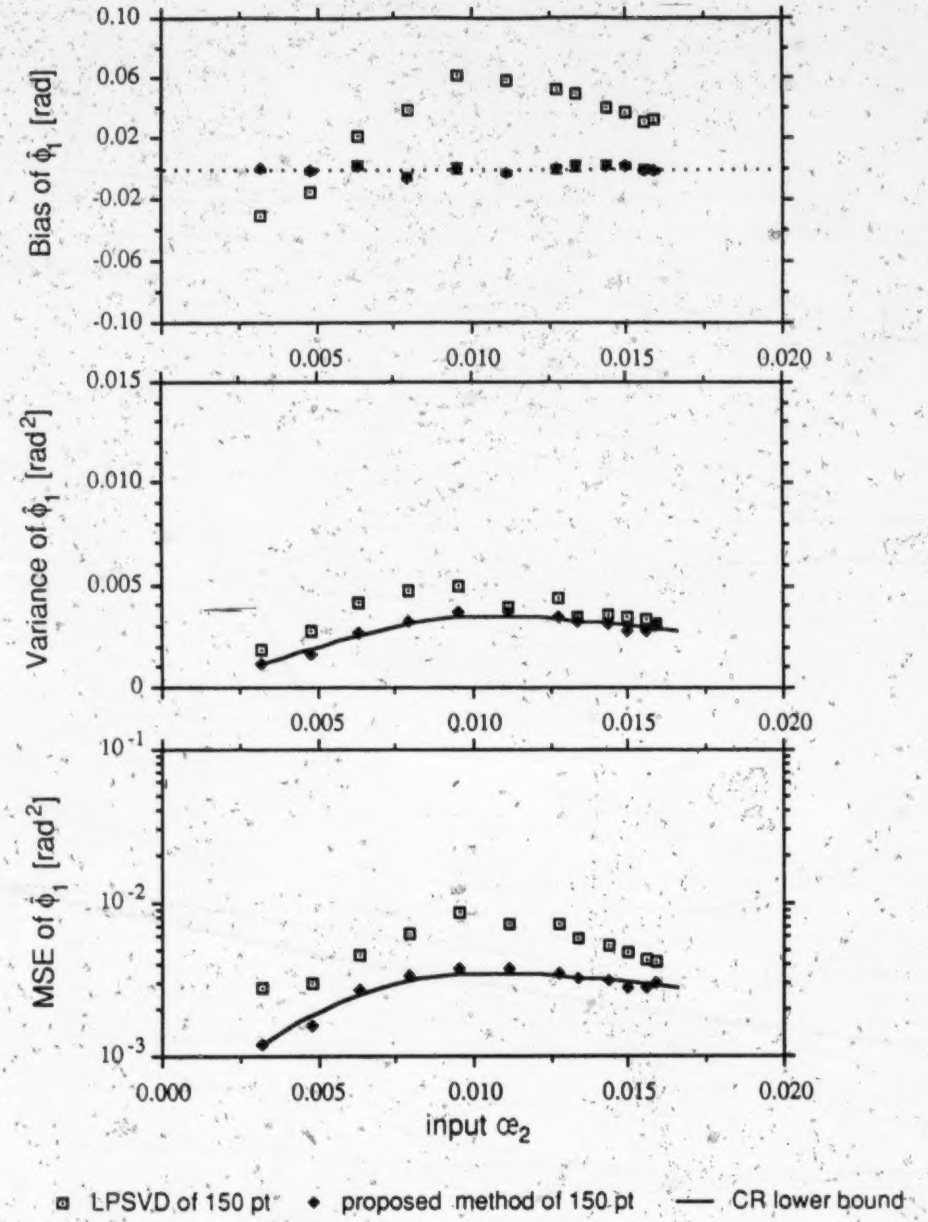


Figure 5.6.1.7 *Phase of the 1-st peak versus input normalized damping rate of the 2-nd peak, œ$_2$. For other information, see Figure 5.6.1.2.*

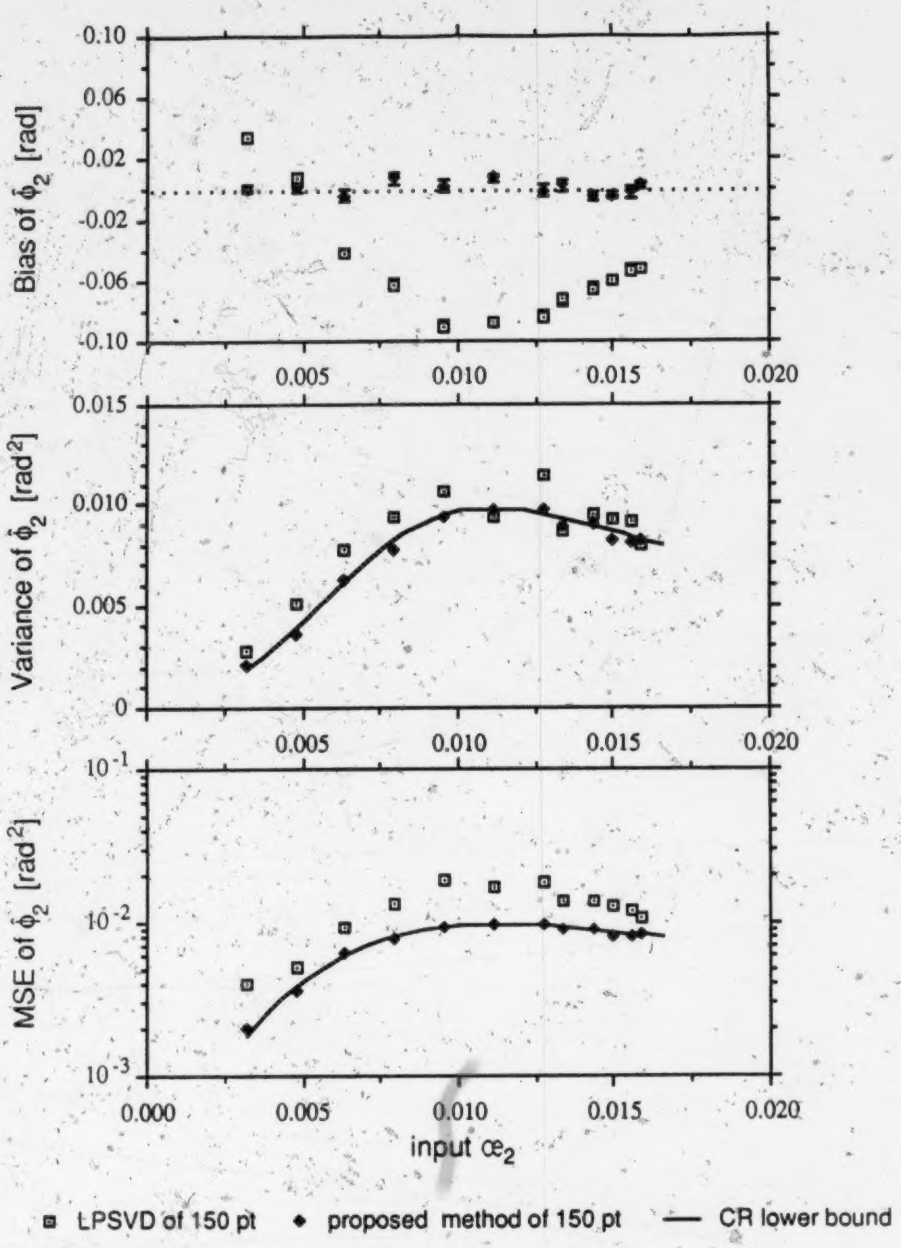


Figure 5.6.1.8 *Phase of the 2-nd peak versus input normalized damping rate of the 2-nd peak, œ_2. For other information, see Figure 5.6.1.2.*

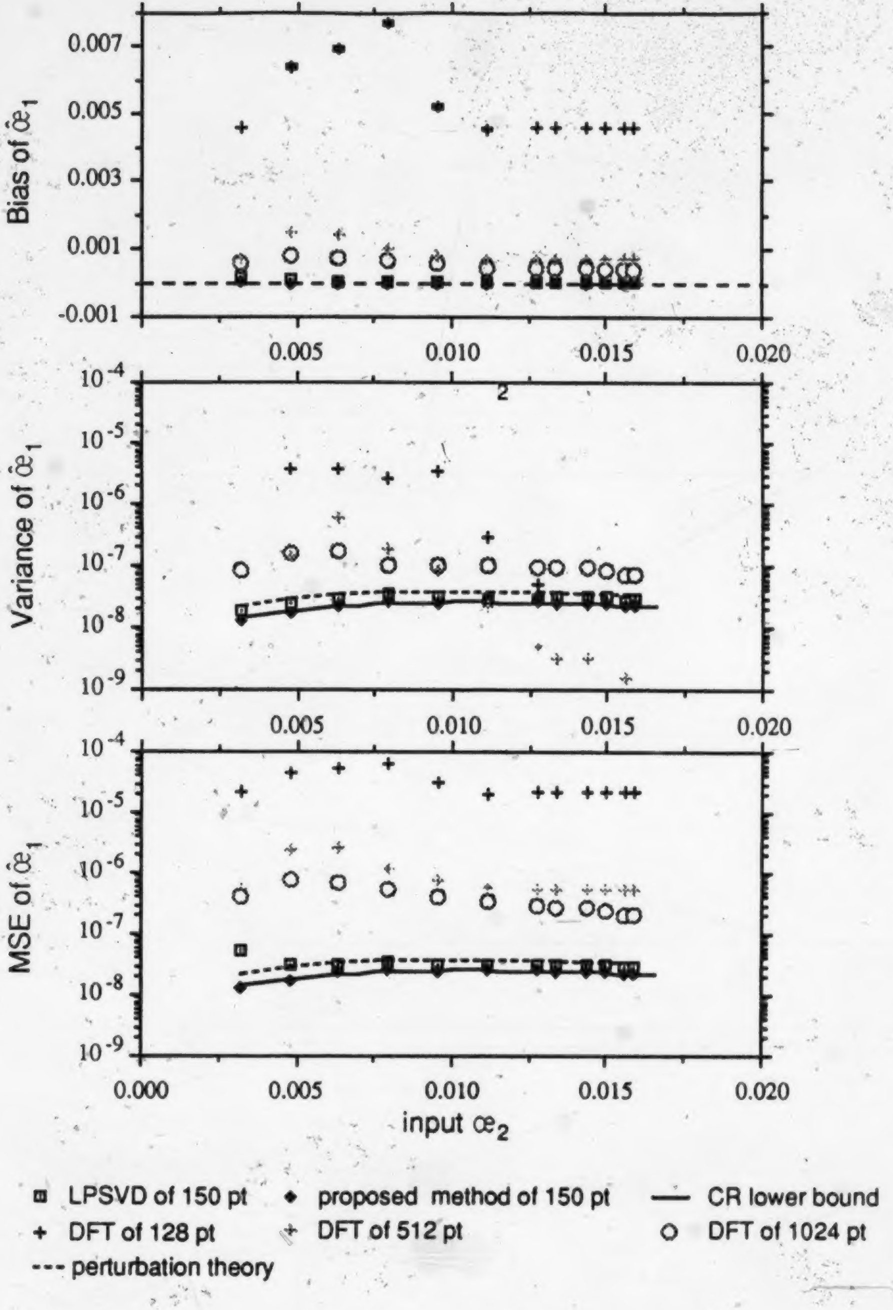


***Figure 5.6.1.9** Normalized damping rate of the 1-st peak versus input normalized damping rate of the 2-nd peak, $œ_2$. For other information, see Figure 5.6.1.2.*

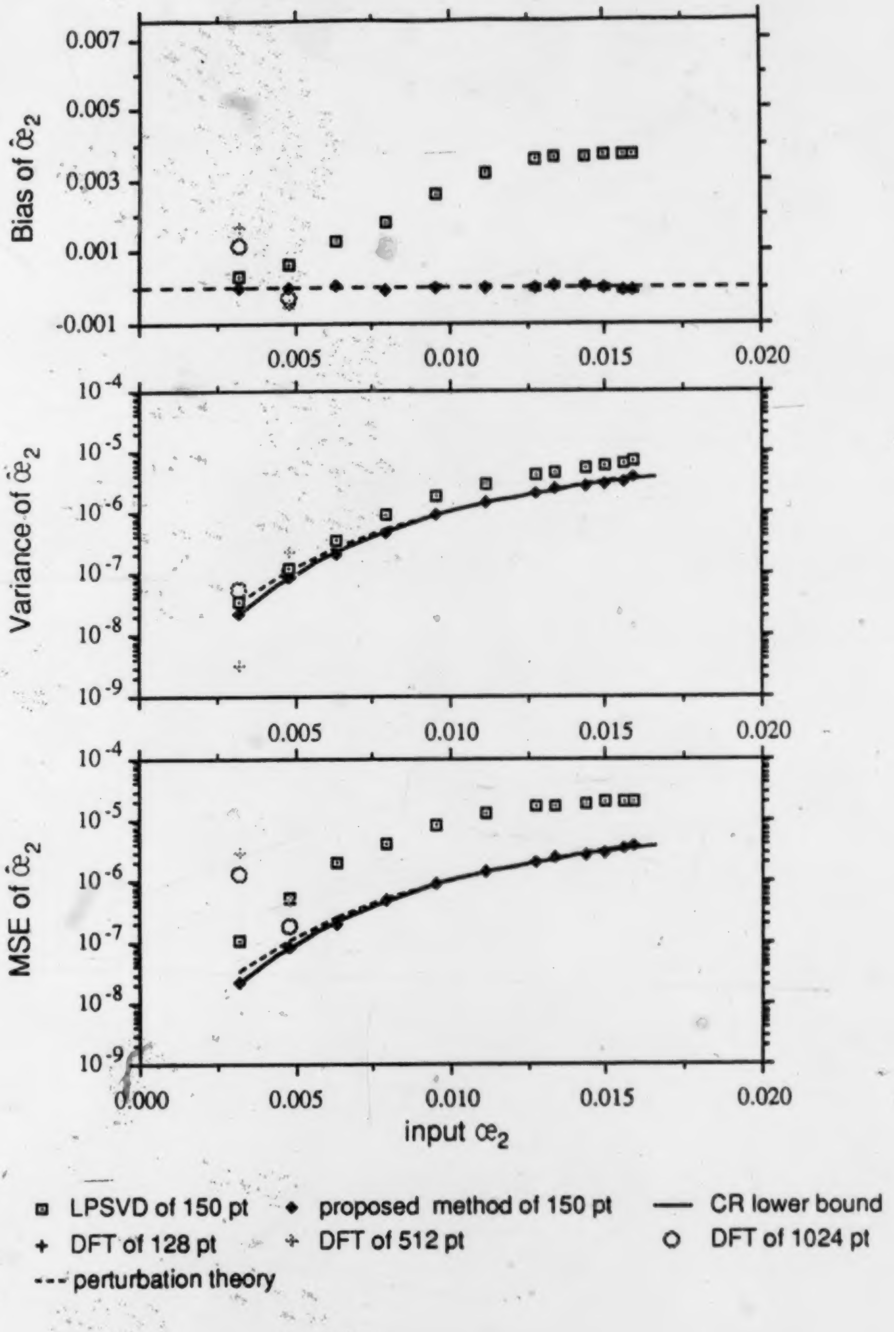


***Figure 5.6.1.10** Normalized damping rate of the 2-nd peak versus input normalized damping rate of the 2-nd peak. For other information, see Figure 5.6.1.2.*

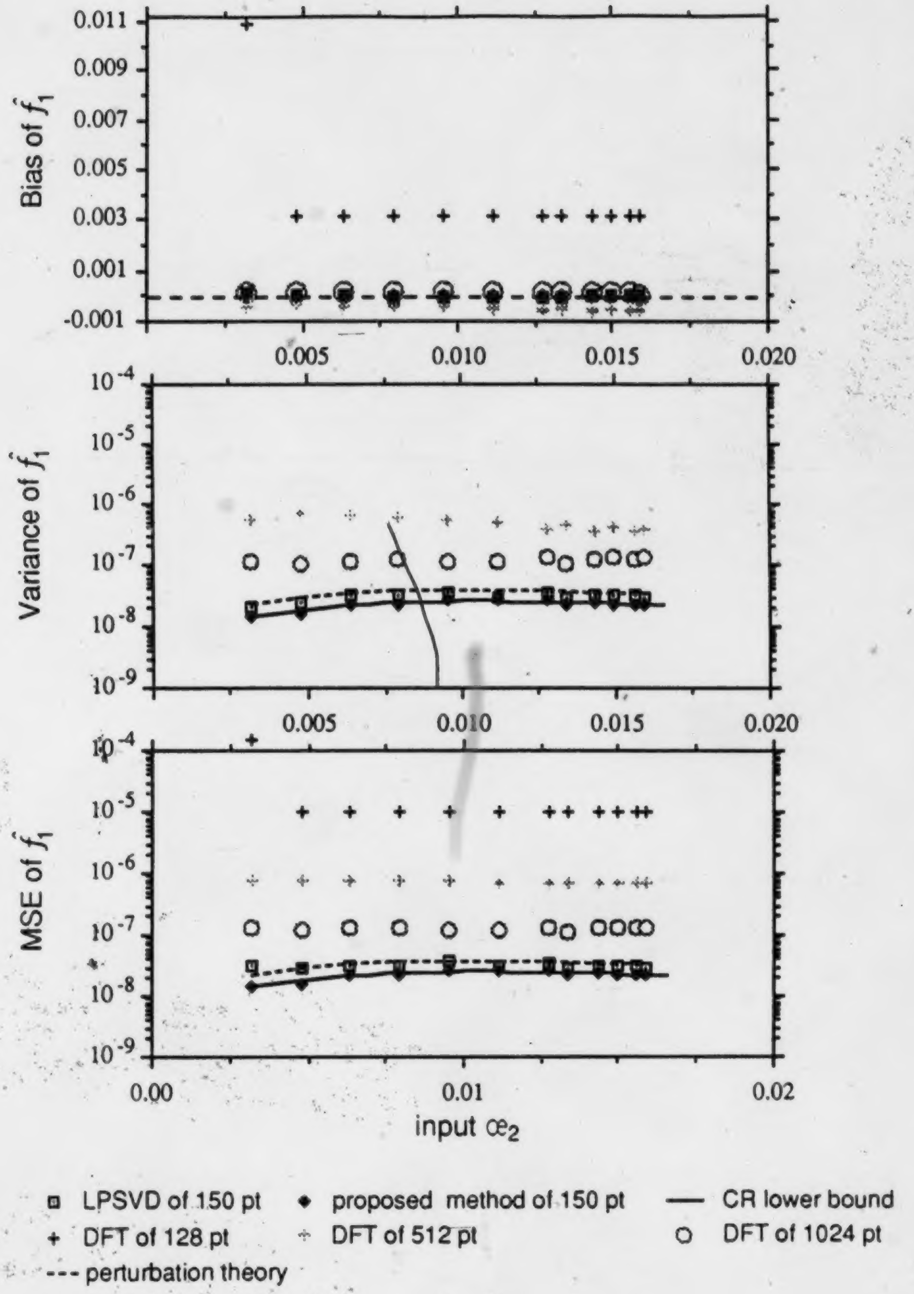


Figure 5.6.1.11 Normalized frequency of the 1-st peak versus input normalized damping rate of the 2-nd peak, $œ_2$. For other information, see Figure 5.6.1.2.

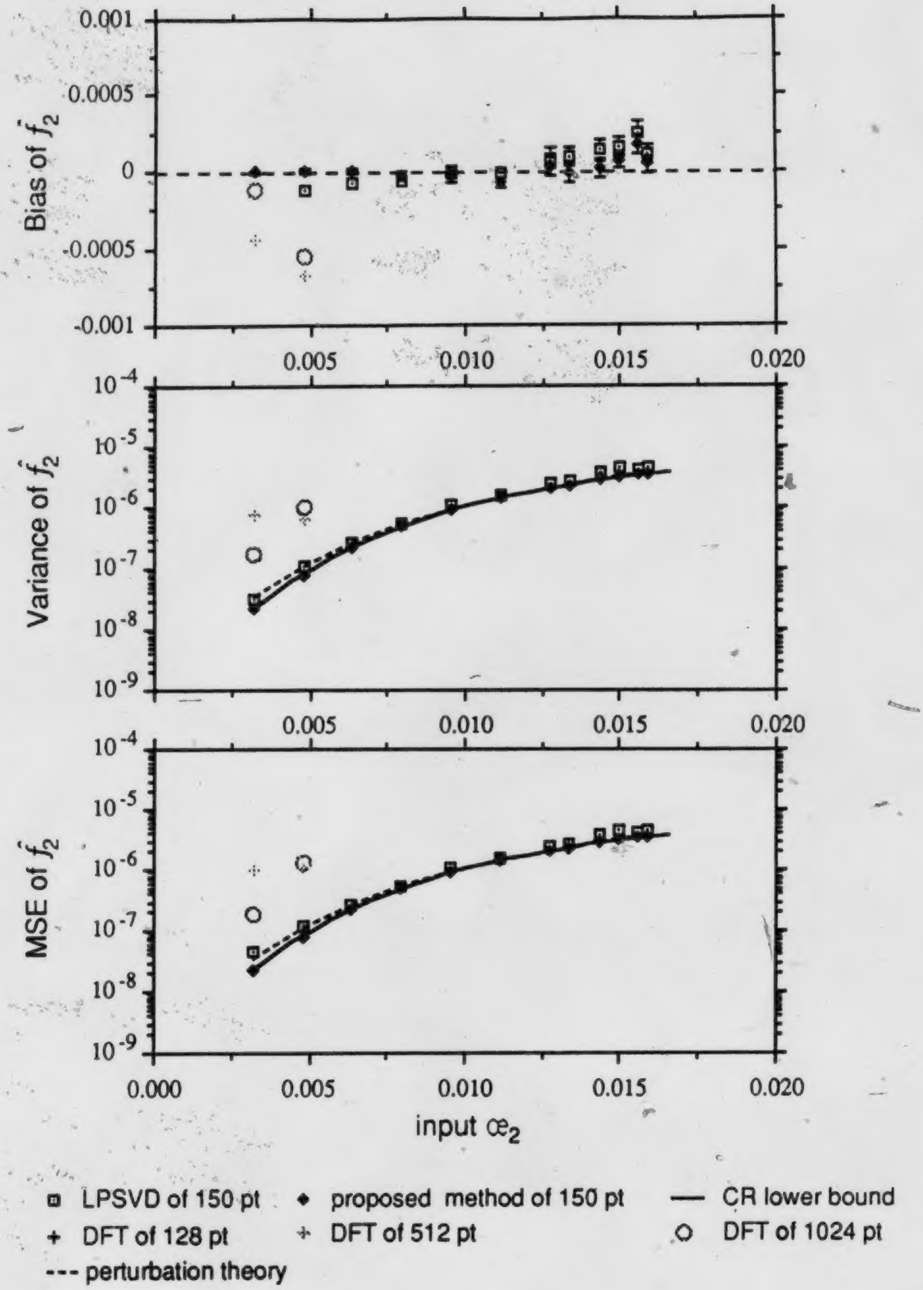


***Figure 5.6.1.12** Normalized frequency of the 2-nd peak versus input normalized damping rate of the 2-nd peak, α_2. For other information, see Figure 5.6.1.2.*

and the proposed method with 150 points can still resolve the peaks for œ_2 increases to 0.016.

Except for the normalized damping rate and frequency of the second peak, the variance estimates of the other parameters from the LPSVD method and the proposed method seem to increase and then gradually decrease again as the input damping rate of the second peak increases. For normalized damping rate œ_2 and normalized frequency f_2, the variances continue to increase as the linewidth (or damping rate) of the second peak increases.

5.7 Effect of changing a peak's amplitude

In this section, we will present an example that evaluates how changes in a component's amplitude affect the performance of the LPSVD method, the proposed method and the DFT method. As before, we will simulate noise corrupted FID data with two exponentially decaying components. All parameters are kept constant, except for the amplitude of the second component. From property 1 of the first order perturbation analysis of the LPSVD method in section 4.2.2, we expect that the variance of damping rate and frequency estimates of the first peak are not affected by changing the amplitude of the second peak. But the corresponding variance estimates of the second peak will be inversely proportional to the magnitude square of the amplitude of the second peak. Now, we will see if simulation results are consistent with perturbation analysis.

Example 5.7.1 :

As before, we have basically the same two signals as in Example 5.2.1, except that the time domain amplitude of the second peak is changed from 25 to 225. For clarity, the time domain amplitudes A_i, phases ϕ_i, normalized damping rates $œ_i$ and normalized frequencies f_i of the two signals are tabulated as follows:

i	A_i	ϕ_i	$œ_i$	f_i
1	250	0	0.003183	-0.3
2	A_2	0	0.003183	-0.275

where A_2 varies from 25 to 225. The sampling interval Δt is still 4 x 10^{-4} sec. Over six hundred independent sequences of complex white Gaussian noise, each having independent real and imaginary parts of variance $\sigma_w^2 = (12.5)^2 = 156.25$ are added to the noise-free signals for each value of A_2. For the LPSVD method and the proposed method, we used 150 data points. For the DFT method, we used 128, 512 and 1024 points. The tentative model order for the LPSVD method is 75.

Figure 5.7.1.1 shows some representative DFT spectra for different time domain amplitudes of the second peak and transformed from 512 and 1024 data points, respectively.

Figure 5.7.1.2 shows the statistical result of the magnitude of the first eight singular values from the LPSVD method. Figure 5.7.1.3 shows the magnitude of the first three singular values versus input time domain amplitude of the second peak, A_2. It seems that as the input A_2 changes, both the first and second singular

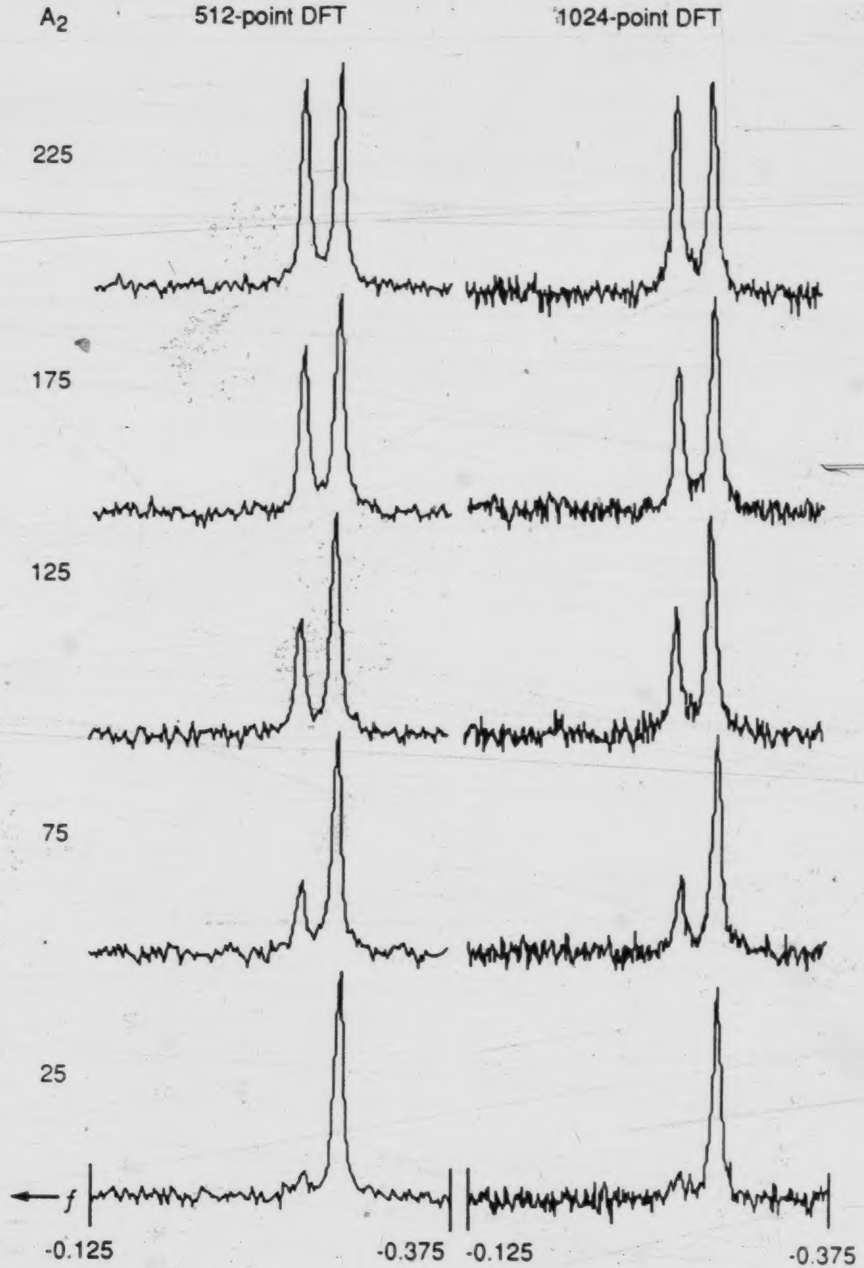


Figure 5.7.1.1 *Some representative DFT spectra for different time domain amplitudes of the 2-nd peak, transformed from 512 and 1024 point FID data which is composed of two decaying sinusoids plus Gaussian noise. The time domain parameters are : amplitude* $A_1 = 250$*; normalized damping rates* $\alpha_1 = \alpha_2 = 0.00318$*; normalized frequencies* $f_1 = -0.3$, $f_2 = -0.275$*; noise standard deviation* σ_w *12.5.*

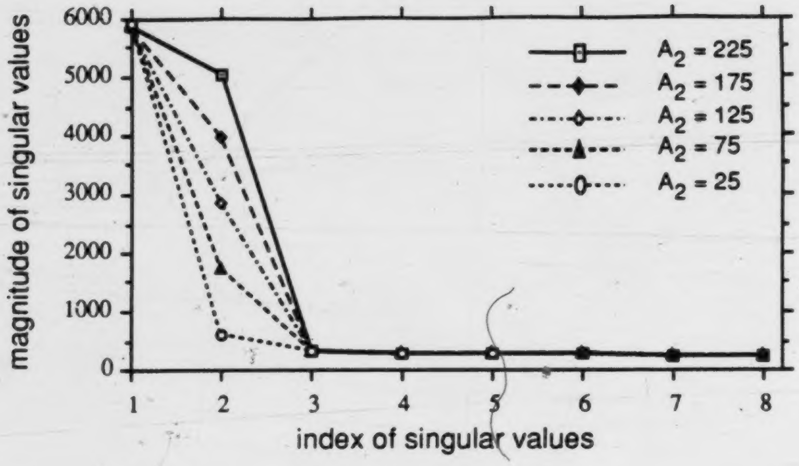


Figure 5.7.1.2 *Statistical result of the first eight singular values of the LPSVD method. A_2 is the input time domain amplitude of the 2-nd peak. Number of data points used is 150. Tentative model order L is 75. For other information, see Figure 5.7.1.1.*

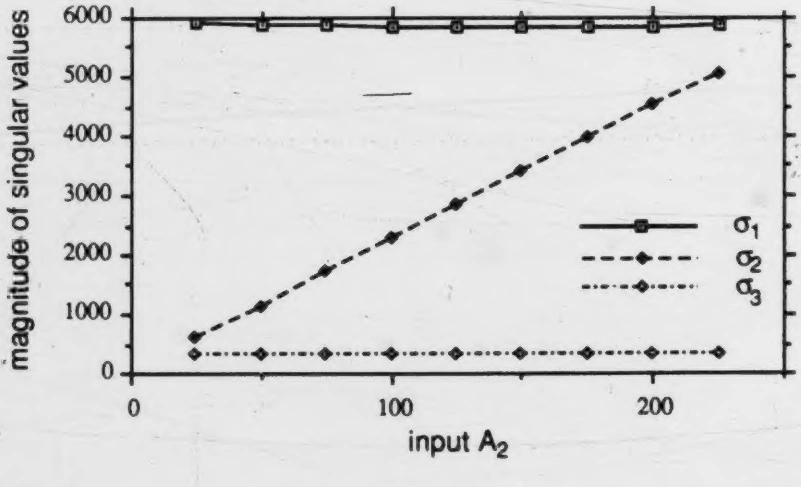


Figure 5.7.1.3 *Magnitude of the first three singular values of the LPSVD method versus input time domain amplitude of the 2-nd peak, A_2. For other information, see Figure 5.7.1.2.*

values change. The largest singular value σ_1 decreases and then increases as input A_2 increases. But the second singular value σ_2 increases linearly as the input A_2 increases. The noise singular values stay constant in the range of input A_2 being studied in this example.

Figures 5.7.1.4 to 5.7.1.12 show the statistical results of the performance measures for the estimates of the spectral parameters from different calculation methods. In general, the proposed method performs better than the LPSVD method, which in turn performs better than the 512-point DFT method, which in turn performs better than the 128-point DFT method. The perturbation analyses from section 4.2.2 are consistent with the simulation results of the LPSVD method for the variance estimate of the damping rates and frequencies. In other words, the damping rate and frequency variance of the first peak are invariant of input A_2, but those of the second peak are inversely proportional to $|A_2|^2$. The LPSVD method is a biased estimator. The bias estimates of the area ratio, the phase and the normalized damping rate of the second peak seem to increase as input A_2 decreases. The proposed method is an unbiased estimator, and it seems to be able to attain the CR lower bounds. For the DFT method, we make the following observations: first, bias estimates of the parameters obtained even for 512 points deviate, in general, more than those of the LPSVD method. Secondly, the variance estimate of the ratio of the peak areas increases with increasing input A_2. The variance estimate of the amplitudes also tend to increase as input A_2 increases. But, the variance estimates of the normalized damping rate and frequency of the second peak have an opposite tendency.

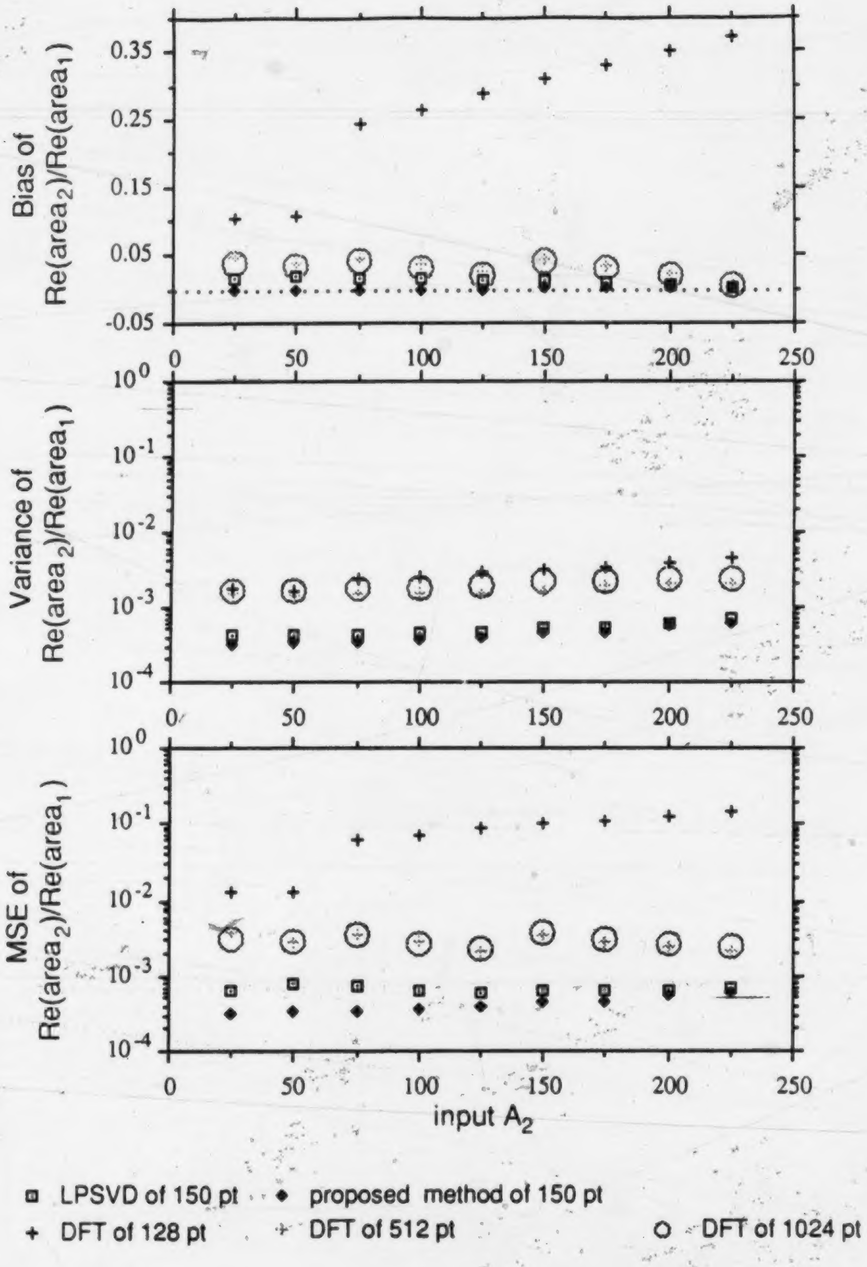


***Figure 5.7.1.4** The area ratio of the two peaks from the real spectrum versus input time domain amplitude of the 2-nd peak, A_2. $\Delta t = 4 \times 10^{-4}$ sec; $A_1 = 250$; $\phi_1 = \phi_2 = 0$; $\alpha_1 = \alpha_2 = 0.00318$. Tentative model order for the LPSVD method is 75. All estimates are obtained from over 600-run simulations.*

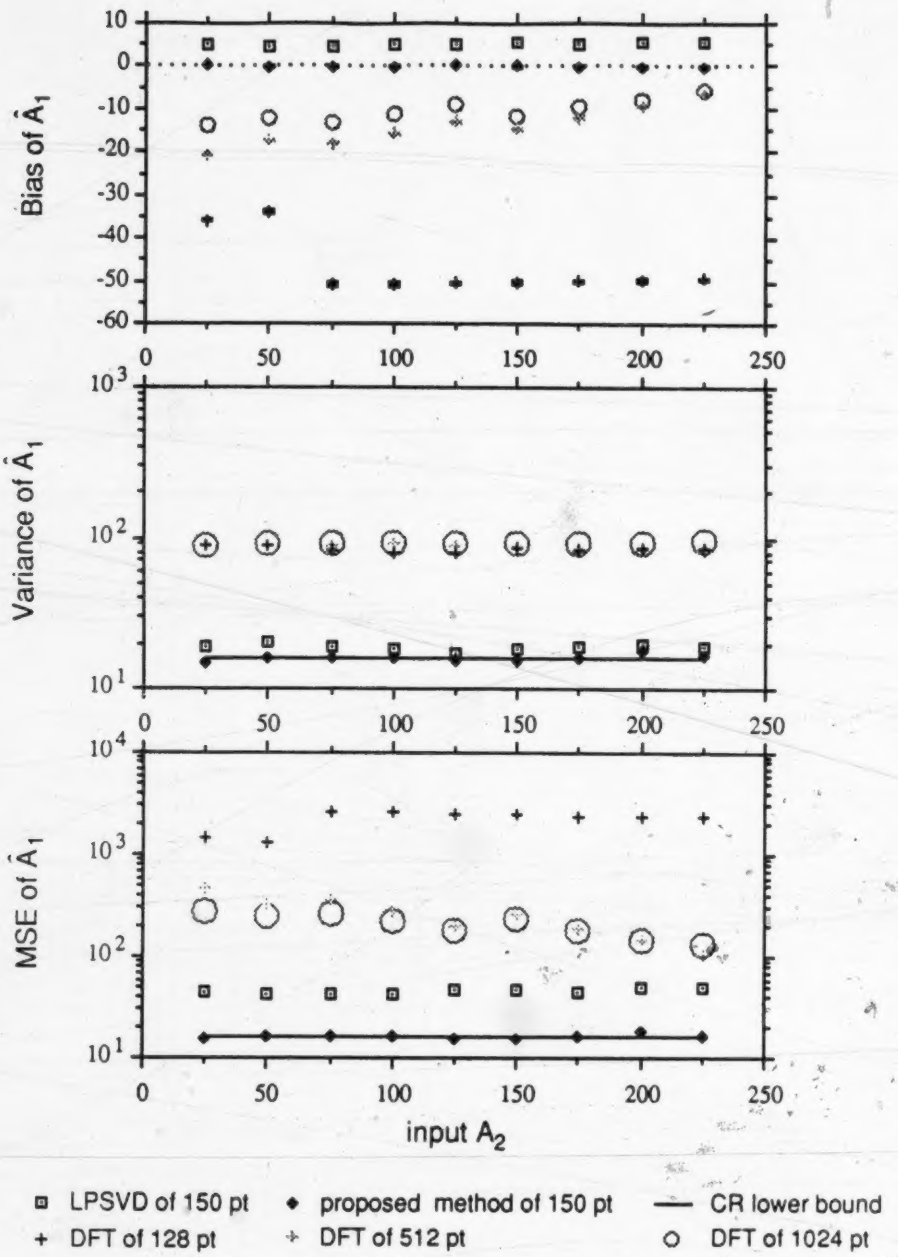


***Figure 5.7.1.5** Time domain amplitude of the 1-st peak versus input A_2. For other information, see Figure 5.7.1.4.*

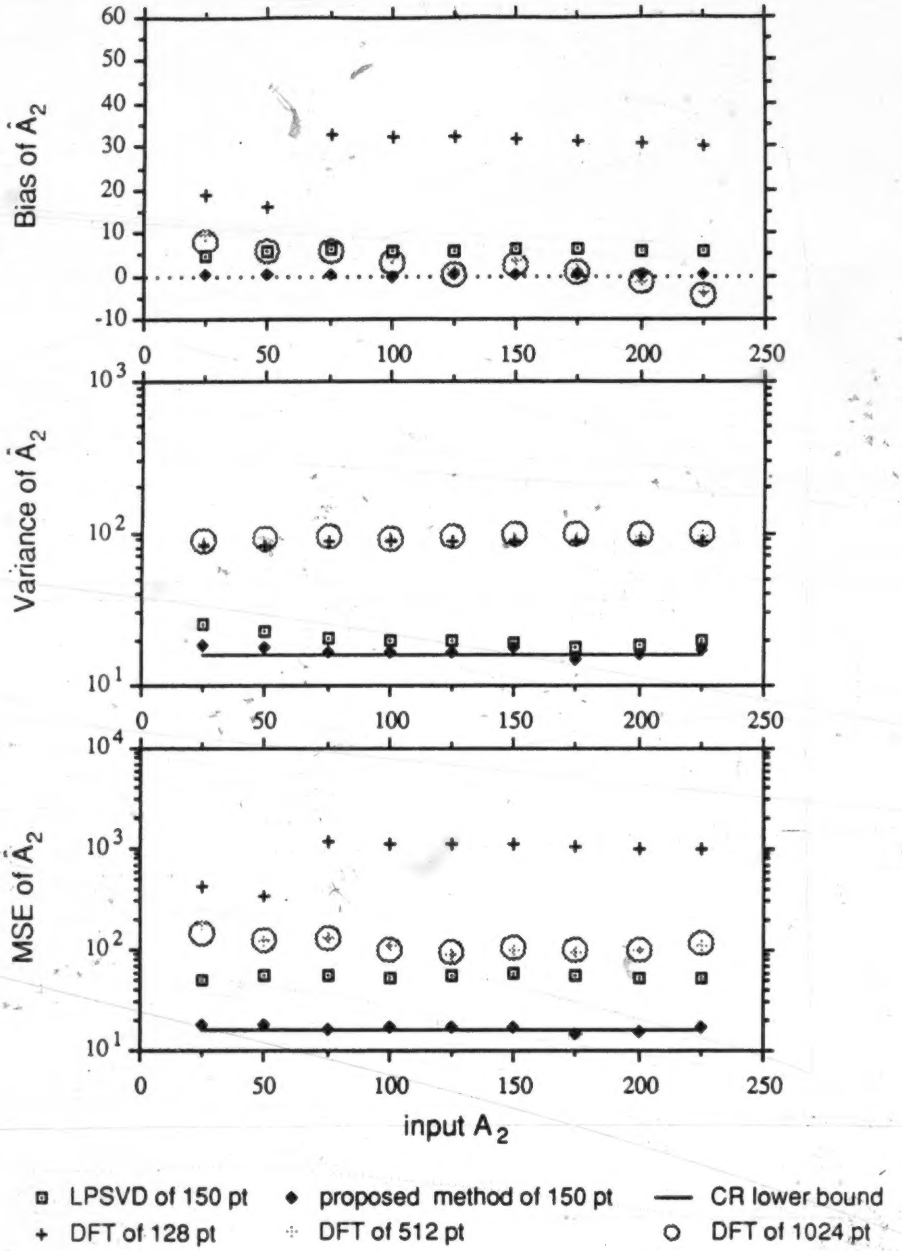


Figure 5.7.1.6 *Time domain amplitude of the 2-nd peak versus input A_2. For other information, see Figure 5.7.1.4.*

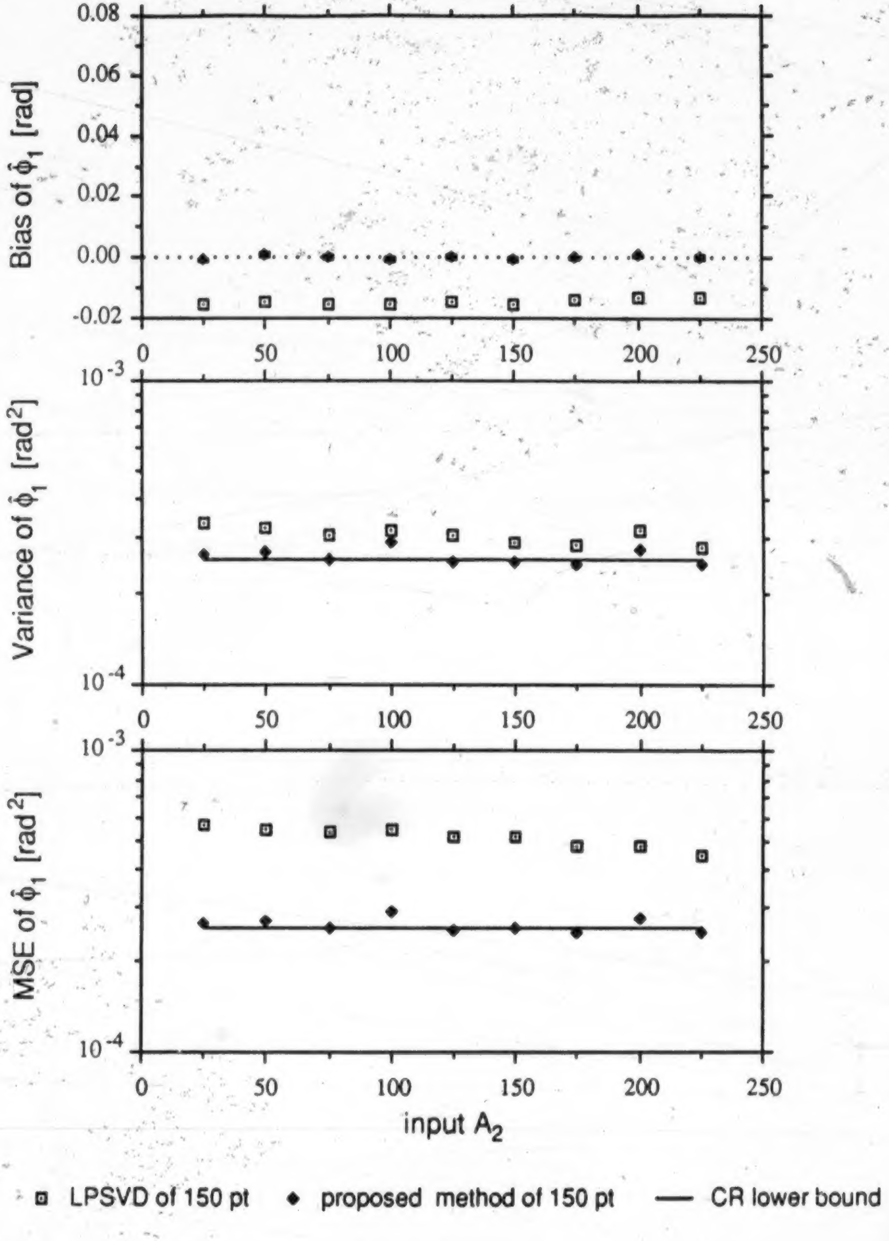


Figure 5.7.1.7 *Phase of the 1-st peak versus input A_2. For other information, see Figure 5.7.1.4.*

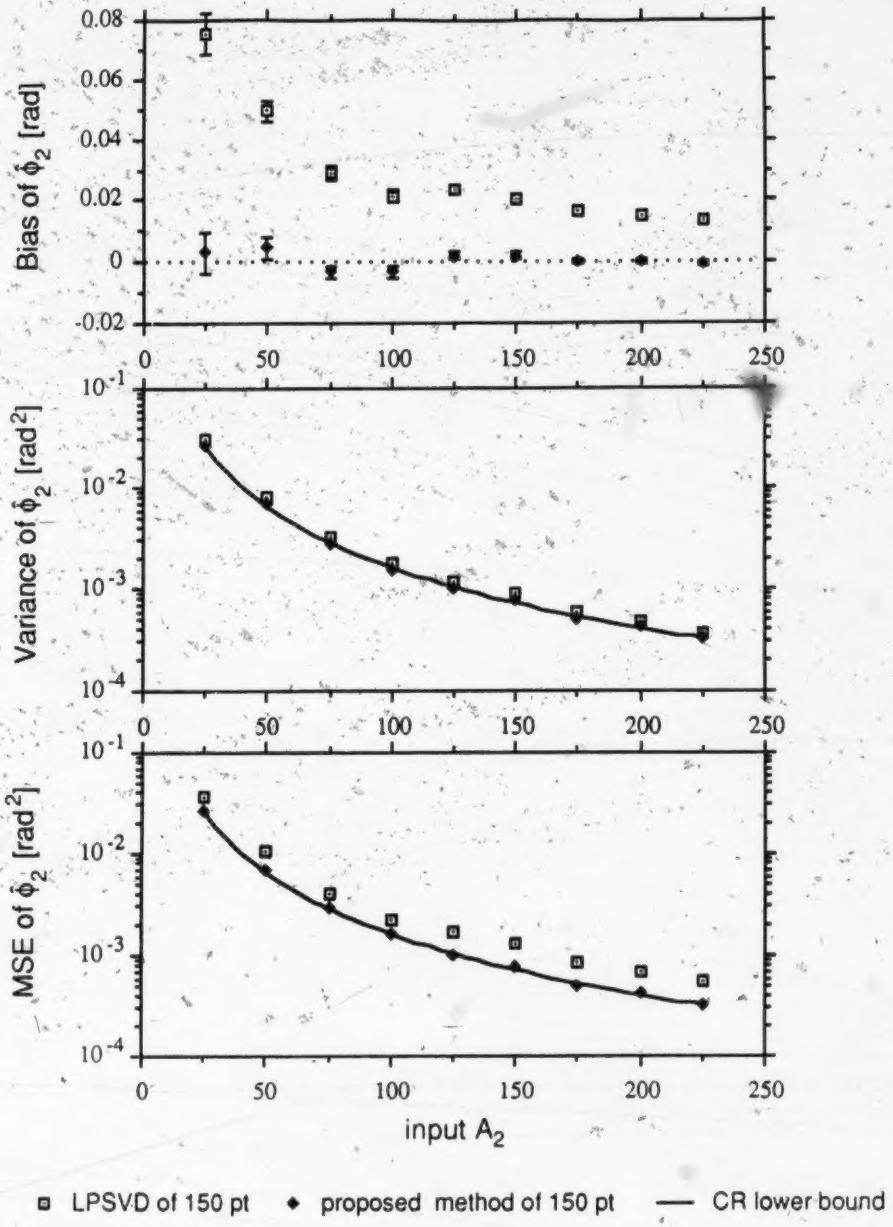


Figure 5.7.1.8 *Phase of the 2-nd peak versus input A_2. For other information, see Figure 5.7.1.4.*

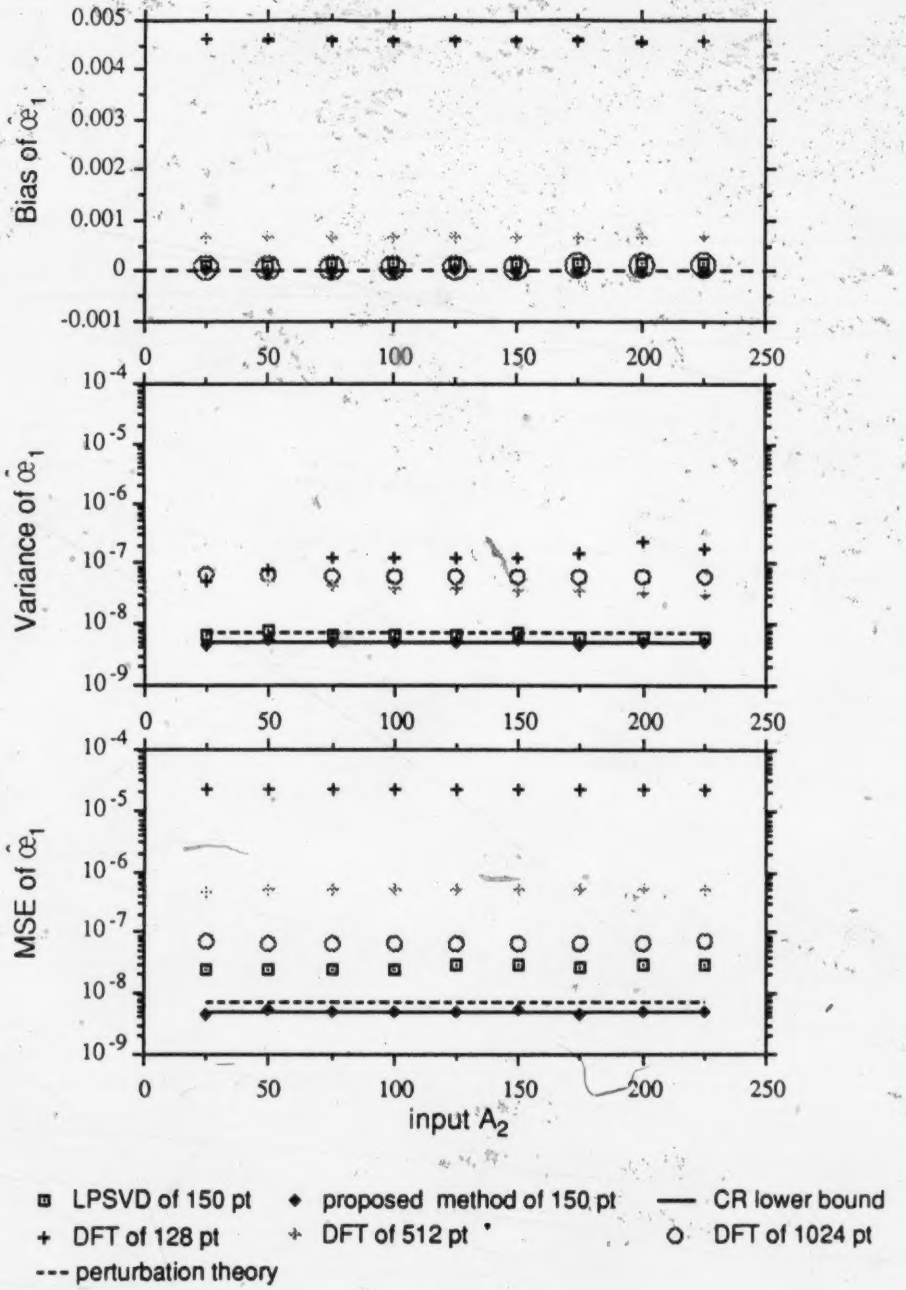


Figure 5.7.1.9 Normalized damping rate of the 1-st peak versus input A_2. For other information, see Figure 5.7.1.4.

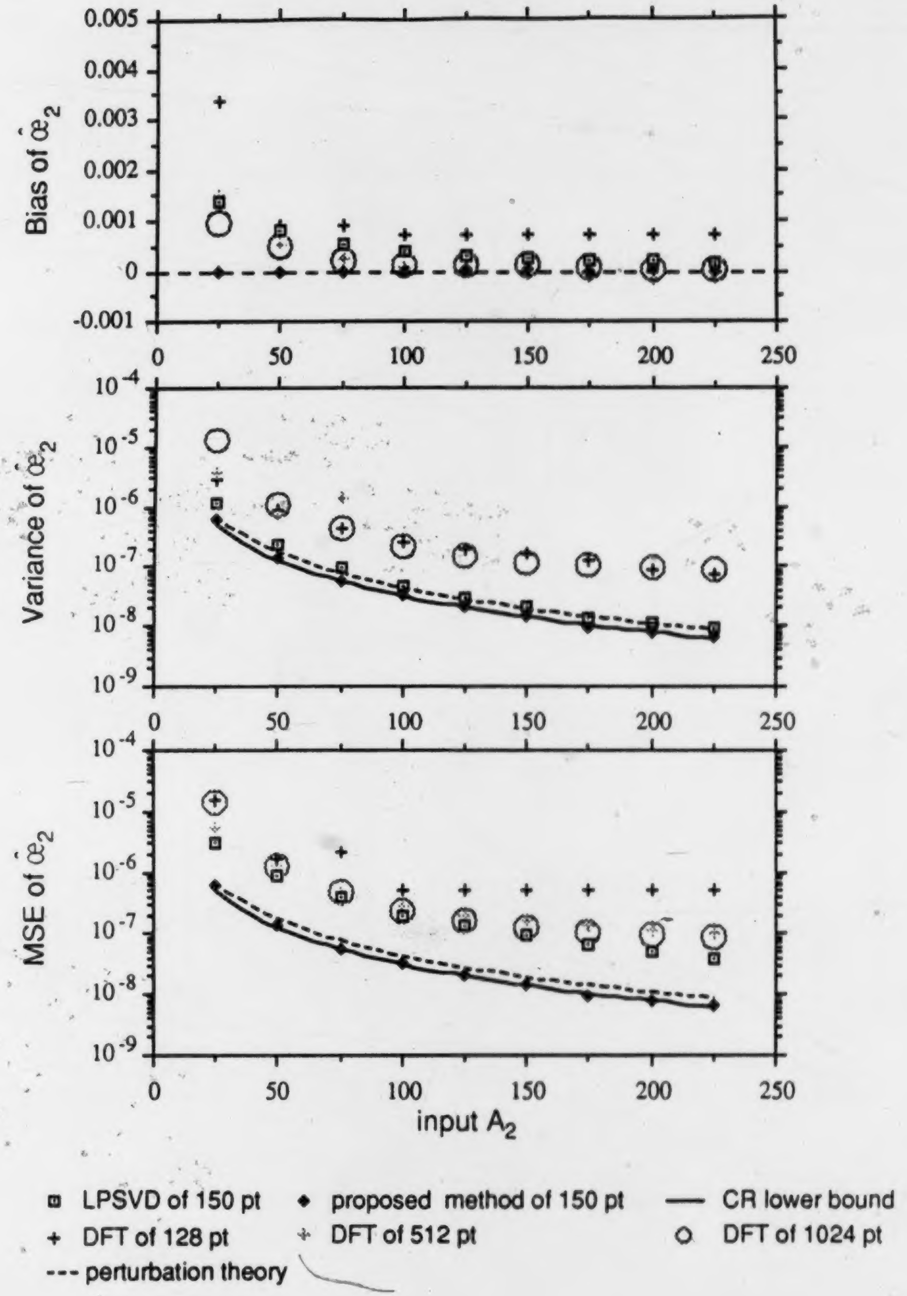


***Figure 5.7.1.10** Normalized damping rate of the 2-nd peak versus input A_2. For other information, see Figure 5.7.1.4.*

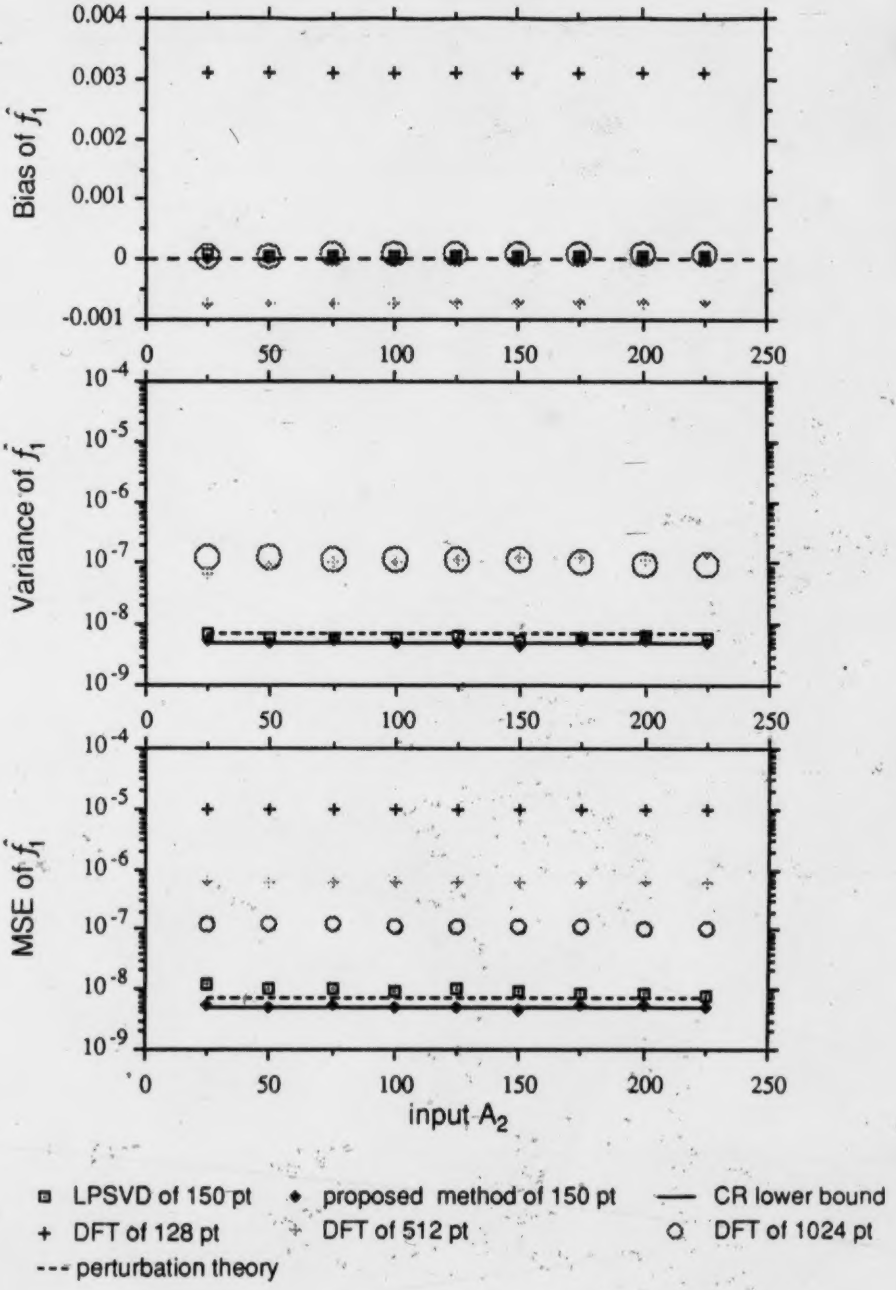


Figure 5.7.1.11 Normalized frequency of the 1-st peak versus input A_2. For other information, see Figure 5.7.1.4.

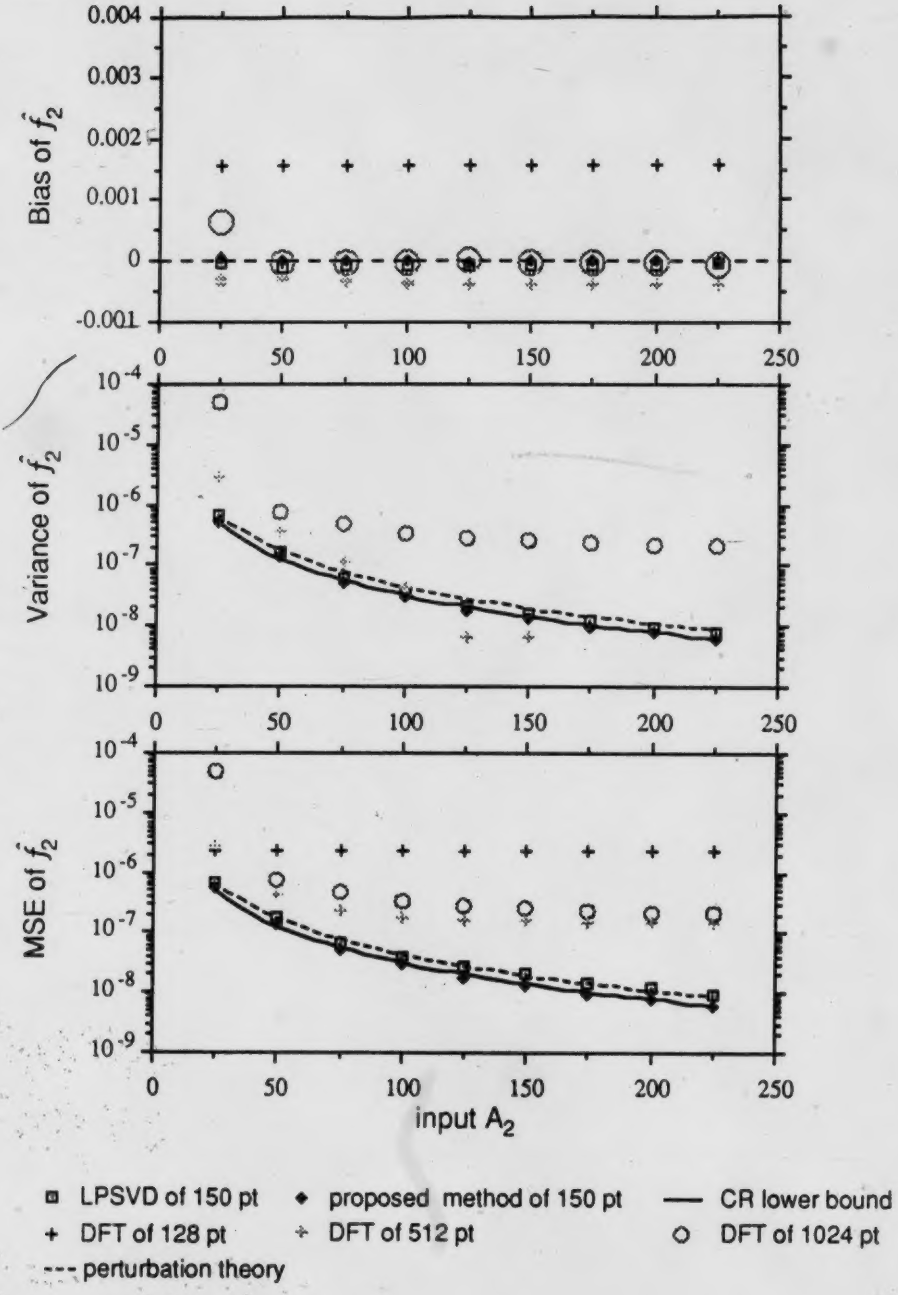


Figure 5.7.1.12 *Normalized frequency of the 2-nd peak versus input A_2. For other information, see Figure 5.7.1.4.*

5.8 Effect of adding a well separated peak

So far, most simulation studies we have carried out consisted of two exponentially decaying components embedded in Gaussian noise. In this subsection, we want to see how much effect the addition of a well separated peak to the spectrum has on the performance of the three methods. We know that if the extra peak is too close to the other two peaks, all methods we are studying will be affected. Peaks that are located too close to one another in the presence of noise can make $\hat{\boldsymbol{A}}_B$ defined by Equation (4.2.1) so ill-conditioned that SVD might not be able to separate the signal singular values from the noise singular values, hence making the LPSVD algorithm fail to estimate the signal parameters accurately. On the other hand, resolution of the DFT method depends on the number of data points being sampled and the linewidth of the peaks. Even if the peaks can be seen on the spectrum, highly damped signals located close to one another are likely to result in a less accurate estimate of the signal parameters. Here, we want to see the effect of adding a well separated narrow peak to the spectrum. The following example is basically the same as Example 5.7.1, except that a third narrow peak located far away from the previous two peaks is added to the spectrum. Other parameters are not changed.

Example 5.8.1 :

The time domain parameters: amplitudes A_i, phases ϕ_i, normalized damping rates œ_i and normalized frequencies f_i, of the noise-free signals are as follows:

i	A_i	ϕ_i	œ_i	f_i
1	250	0	0.003183	-0.3
2	A_2	0	0.003183	-0.275
3	250	0	0.003183	0.1

where A_2 varies from 25 to 225. The sampling interval Δt is still 4×10^{-4} sec. As before, over six hundred independent sets of complex white Gaussian noise, each having an independent real and imaginary part of variance $\sigma_w^2 = (12.5)^2 = 156.25$ are added to the noise-free signals for each value of A_2. For the LPSVD method and the proposed method, we used 150 data points. The tentative model order for the LPSVD method was 75. For the DFT method, we used 128, 512 and 1024 points.

Figure 5.8.1.1 shows some representative DFT spectra for different input values of A_2 transformed from 512 data points.

Figure 5.8.1.2 shows the statistical result of the magnitude of the first six singular values from the LPSVD method. Figure 5.8.1.3 shows the magnitude of the first four singular values versus input time domain amplitude of the third peak, A_2. Comparing with Figure 5.7.1.3, one observes that a signal singular value of magnitude of about 6000 is added to Figure 5.8.1.3. The third singular value in this example is similar to the second singular value in the previous example, and so on.

Figures 5.8.1.4 to 5.8.1.17 show the statistical results of the performance measures for the estimates of the spectral parameters from different calculation methods. Comparing with the results of the previous example, we notice that adding a well separated small decaying component to the previous signals does not change significantly the estimates of the parameters of the first and second components arrived at by the three methods. The estimates of the signal parameters of the third, well separated narrow, peak by the proposed method is better than the LPSVD method, which in turn is better than the DFT method.

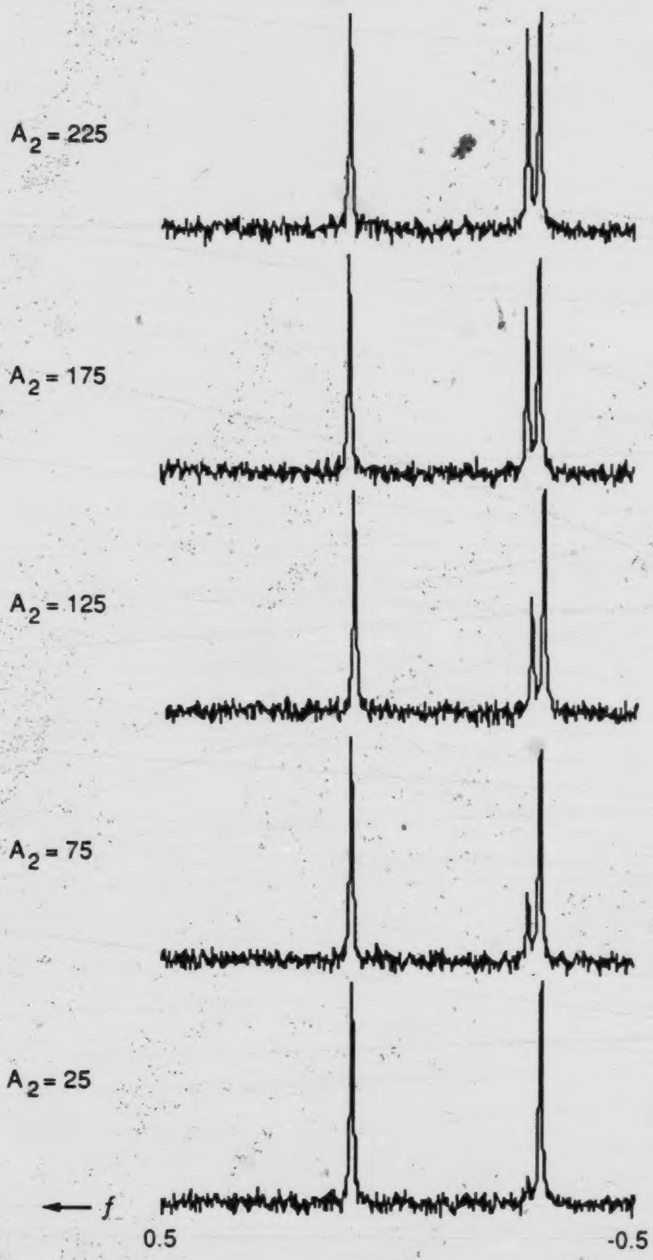


Figure 5.8.1.1 Some representative DFT spectra for different time domain amplitudes of the 2-nd peak, transformed from 512 points FID data, which are composed of three decaying sinusoids plus Gaussian noise. The time domain parameters are : amplitudes $A_1 = A_3 = 250$; normalized damping rates $\alpha_1 = \alpha_2 = \alpha_3 = 0.00318$; normalized frequencies $f_1 = -0.3$, $f_2 = -0.275$, $f_3 = 0.1$; noise standard deviation 12.5.

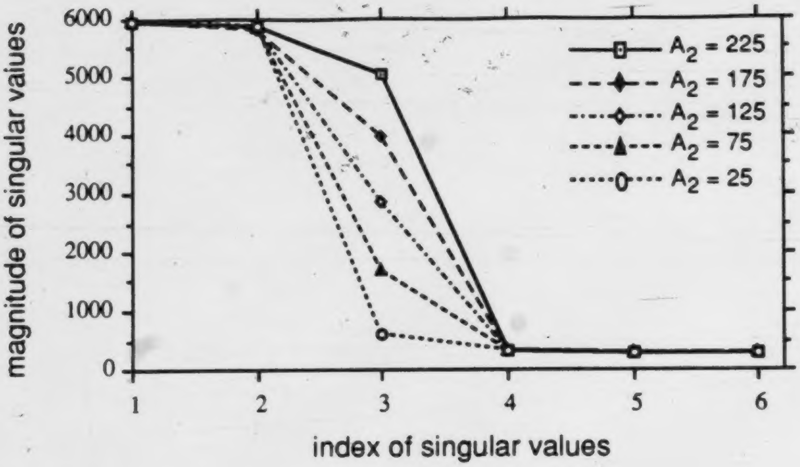


Figure 5.8.1.2 Statistical result of the first six singular values of the LPSVD method. A_2 is the input time domain amplitude of the 2-nd peak. Number of data points used is 150. Tentative model order L is 75. For other information, see Figure 5.8.1.1.

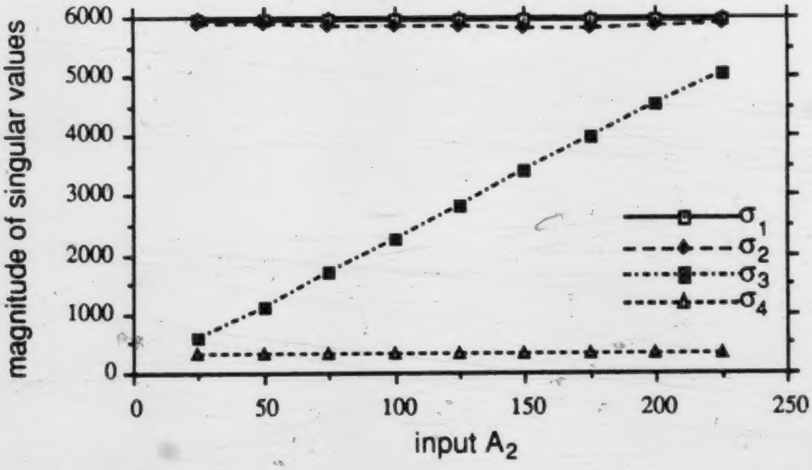


Figure 5.8.1.3 Magnitude of the first four singular values of the LPSVD method versus input time domain amplitude of the 2-nd peak, A_2. For other information, see Figure 5.8.1.2.

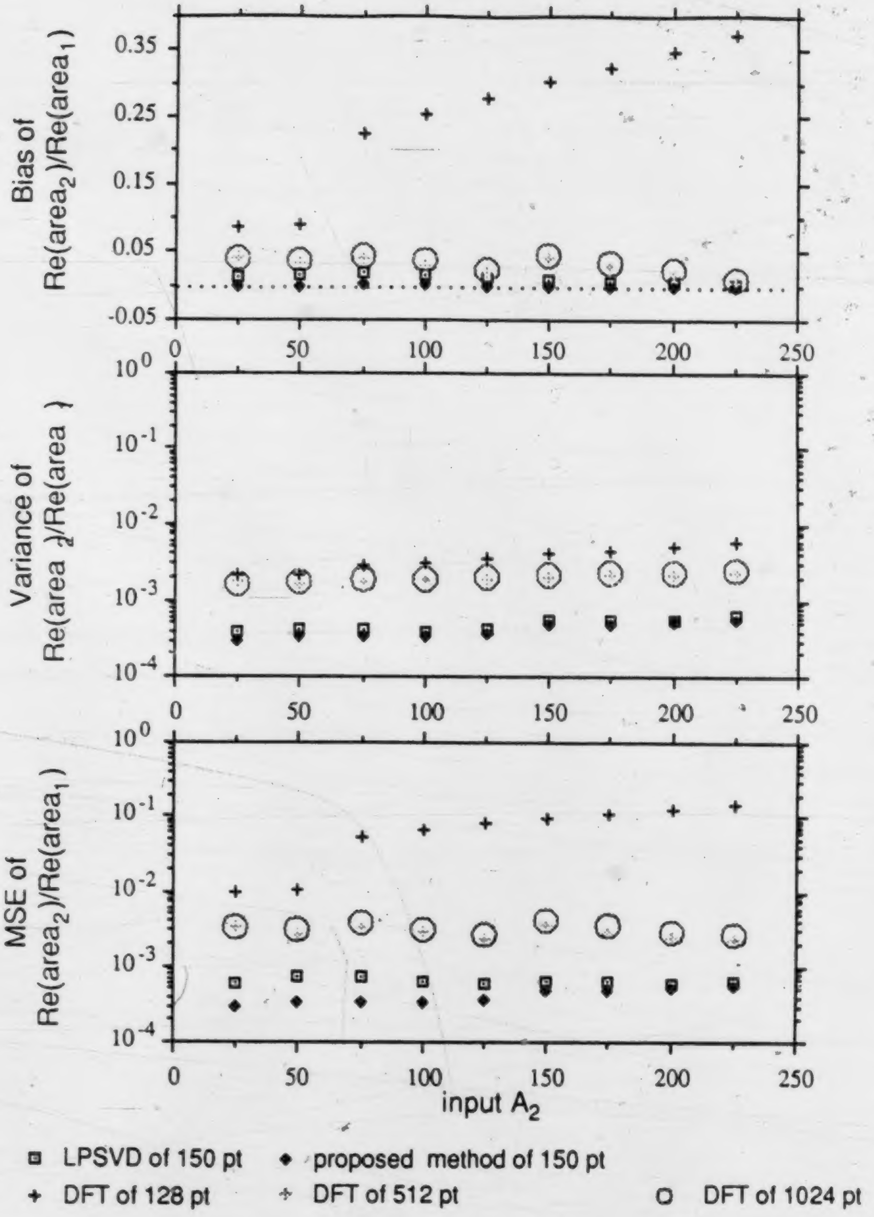


Figure 5.8.1.4 The area ratio of the 2-nd peak to the 1-st peak from the real spectrum versus input time domain amplitude of the 2-nd peak, A_2. Tentative model order for the LPSVD method is 75. All estimates are obtained from over 600-run simulations. For other information, see Figure 5.8.1.1.

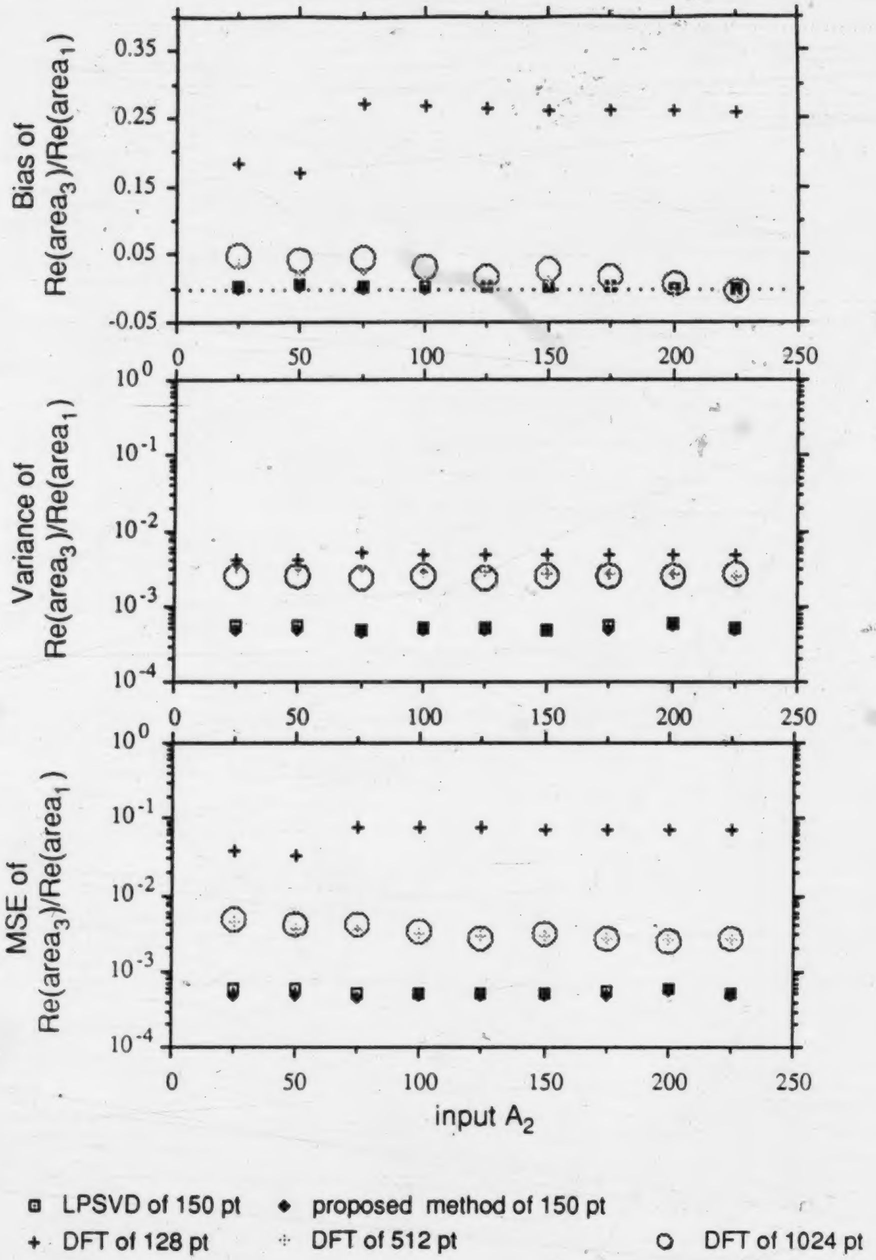


Figure 5.8.1.5 *The area ratio of the 3-rd peak to the 1-st peak from the real spectrum versus input time domain amplitude of the 2-nd peak, A_2. For other information, see Figure 5.8.1.4.*

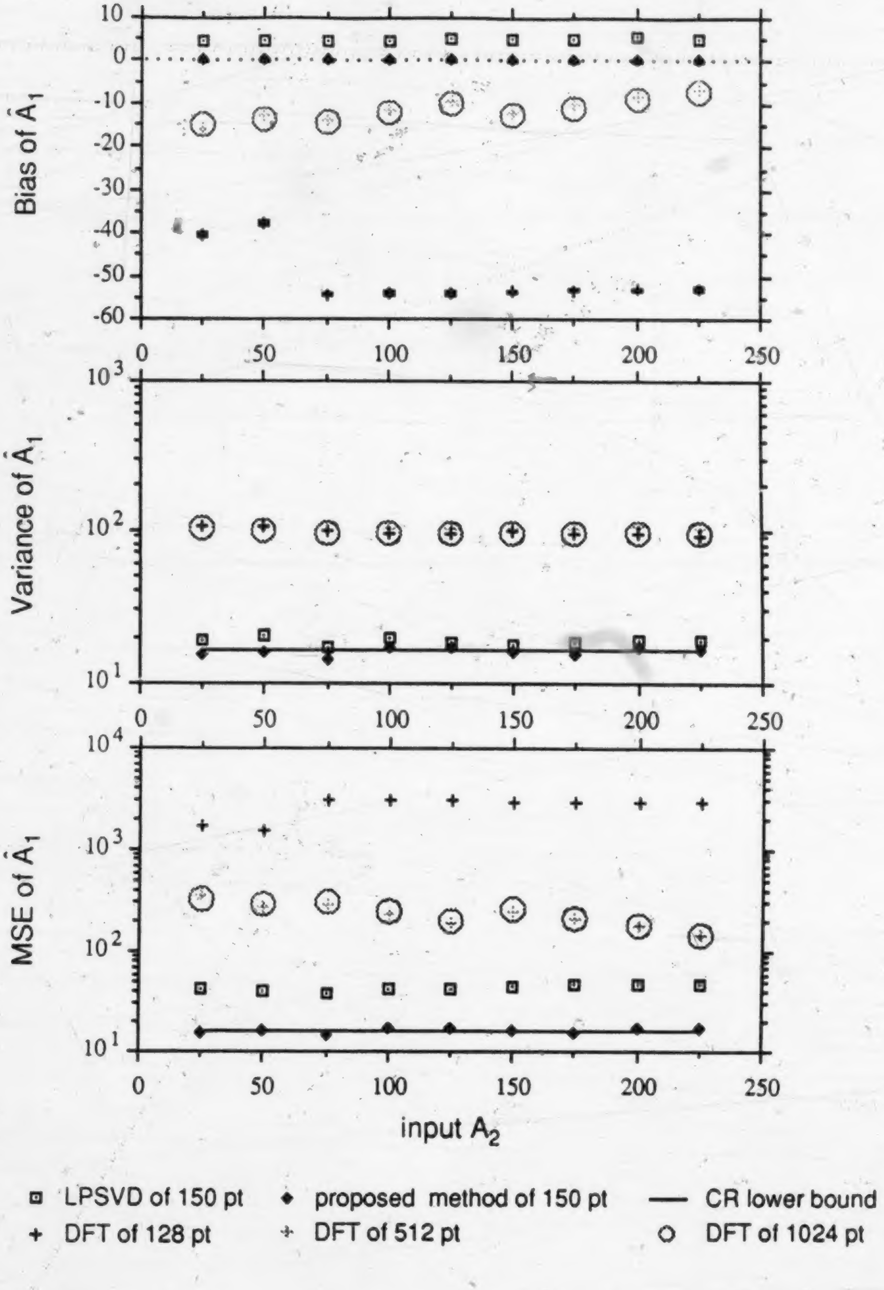


Figure 5.8.1.6 *Time domain amplitude of the 1-st peak versus input,* A_2*. For other information, see Figure 5.8.1.4.*

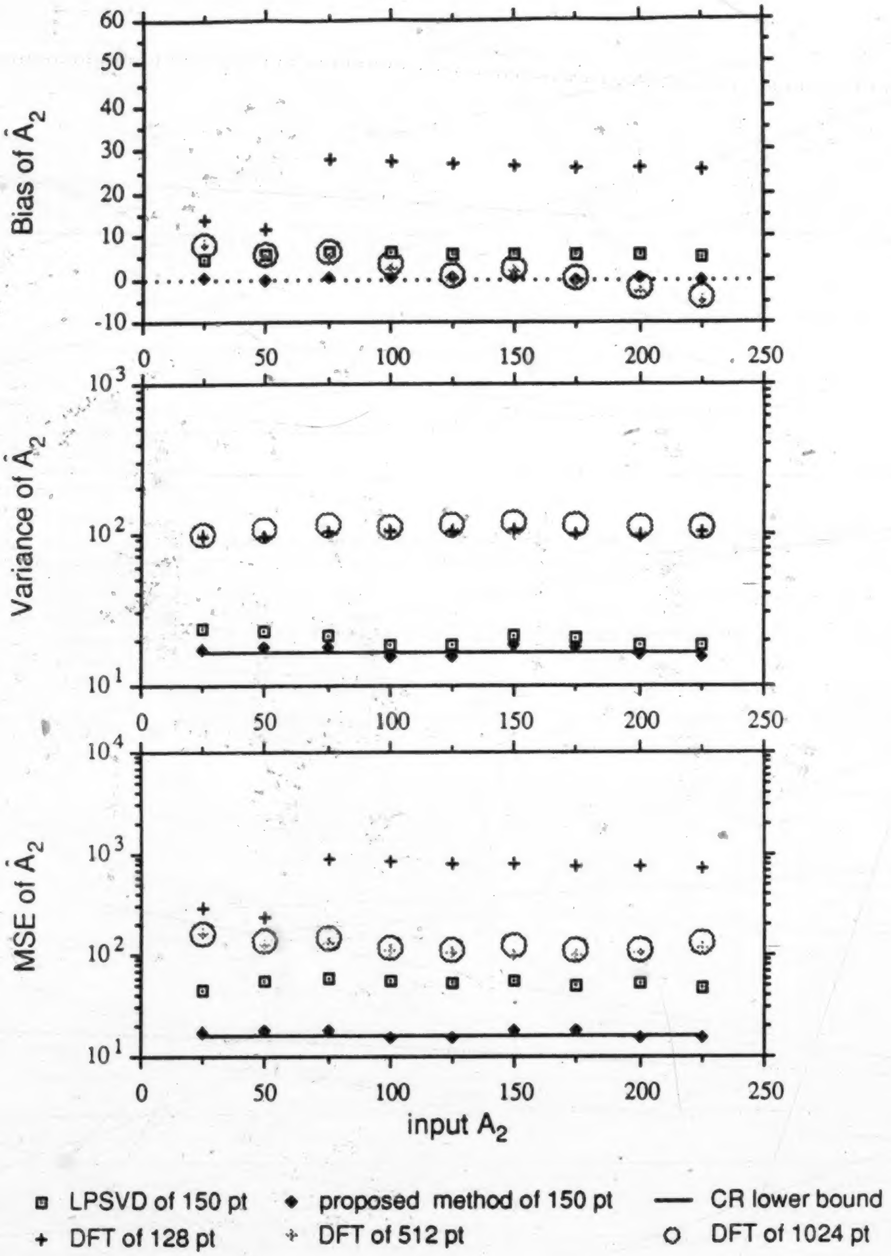


Figure 5.8.1.7 Time domain amplitude of the 2-nd peak versus input A_2. For other information, see Figure 5.8.1.4.

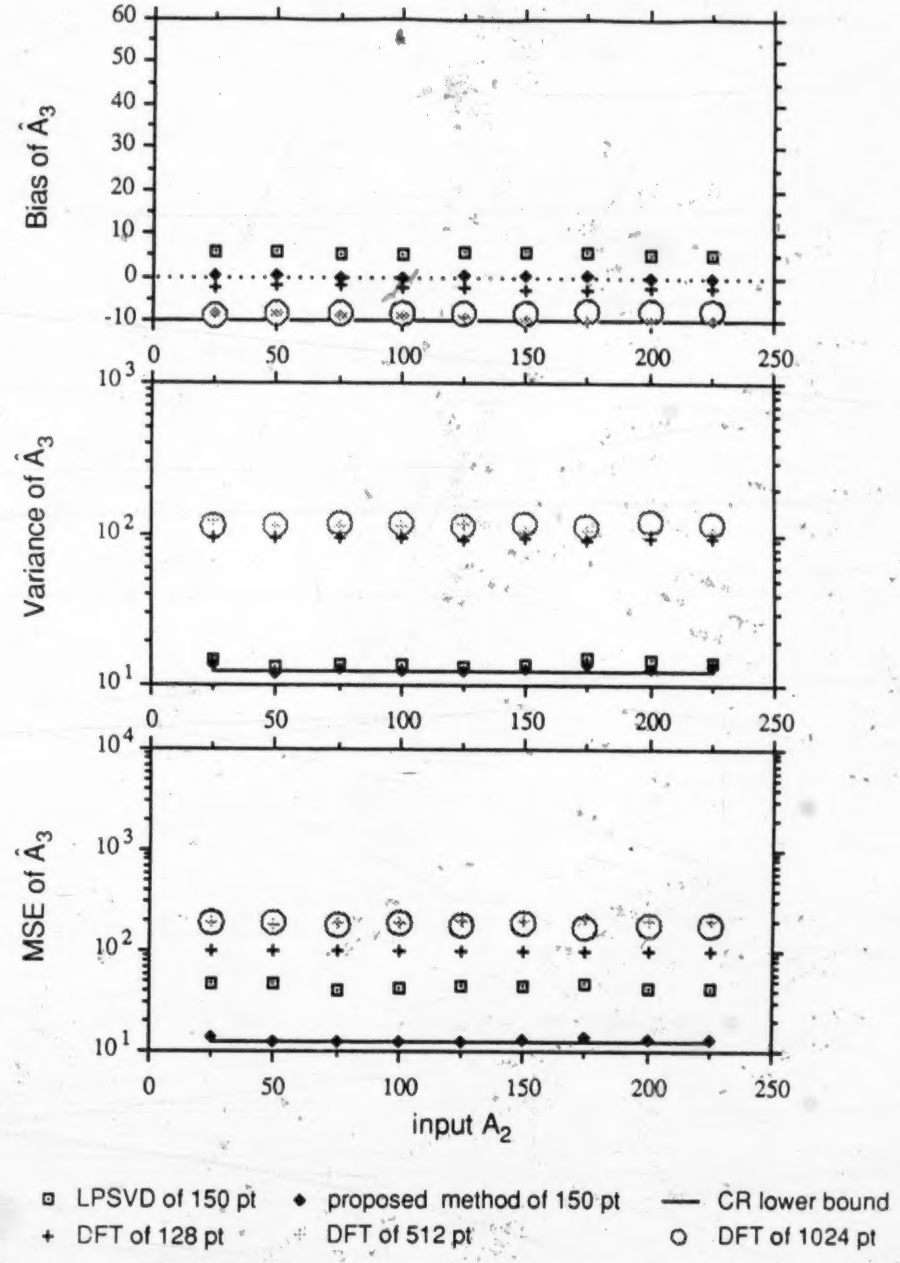


Figure 5.8.1.8 *Time domain amplitude of the 3-rd peak versus input A_2. For other information, see Figure 5.8.1.4.*

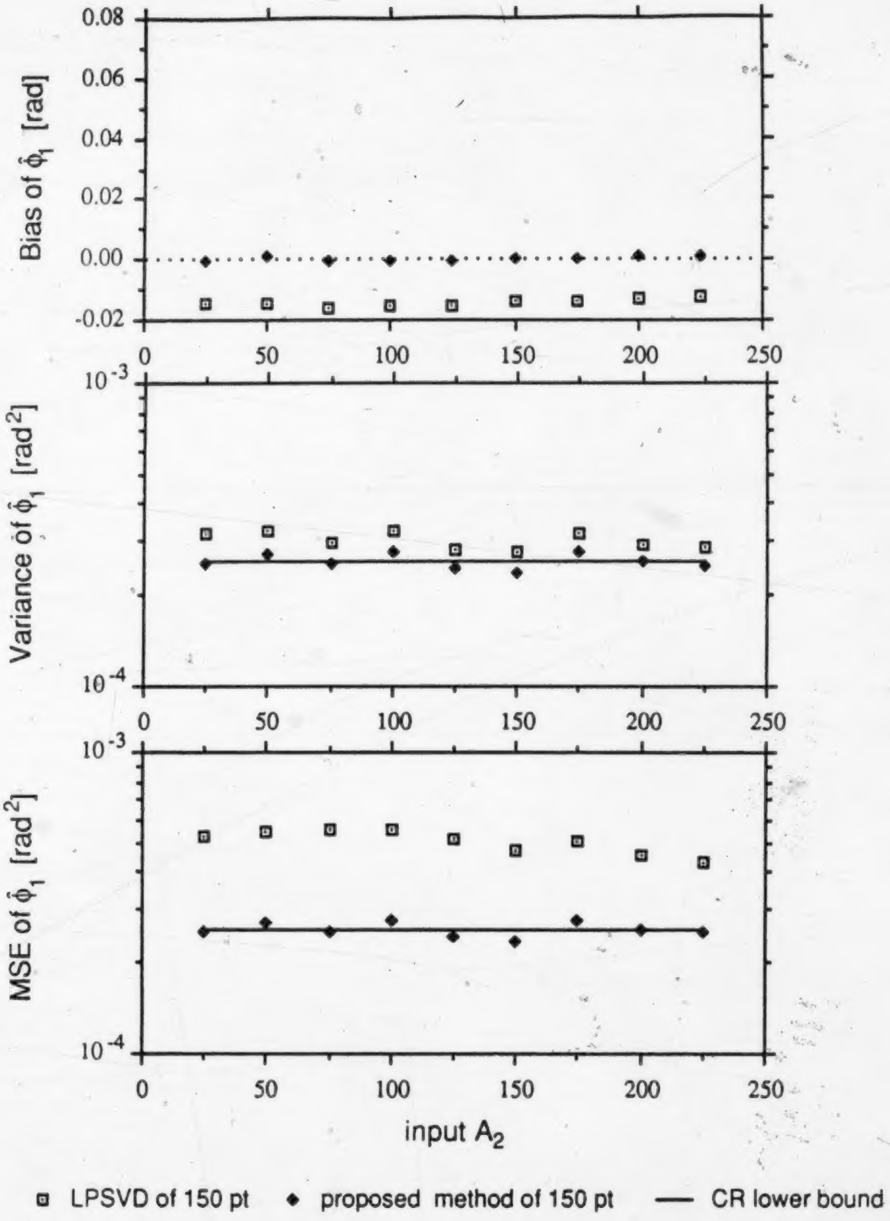


***Figure 5.8.1.9** Phase of the 1-st peak versus input A_2. For other information, see Figure 5.8.1.4.*

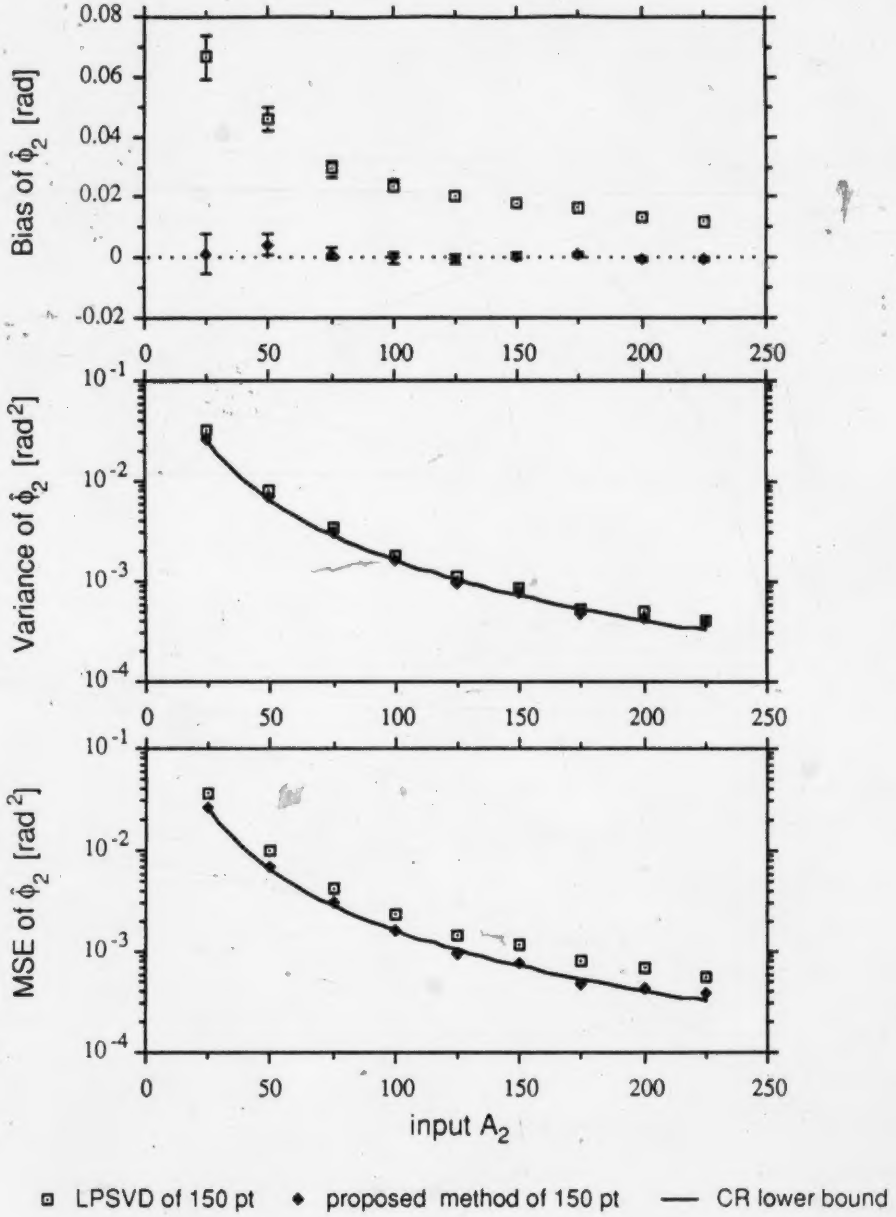


Figure 5.8.1.10 *Phase of the 2-nd peak versus input A_2. For other information, see Figure 5.8.1.4.*

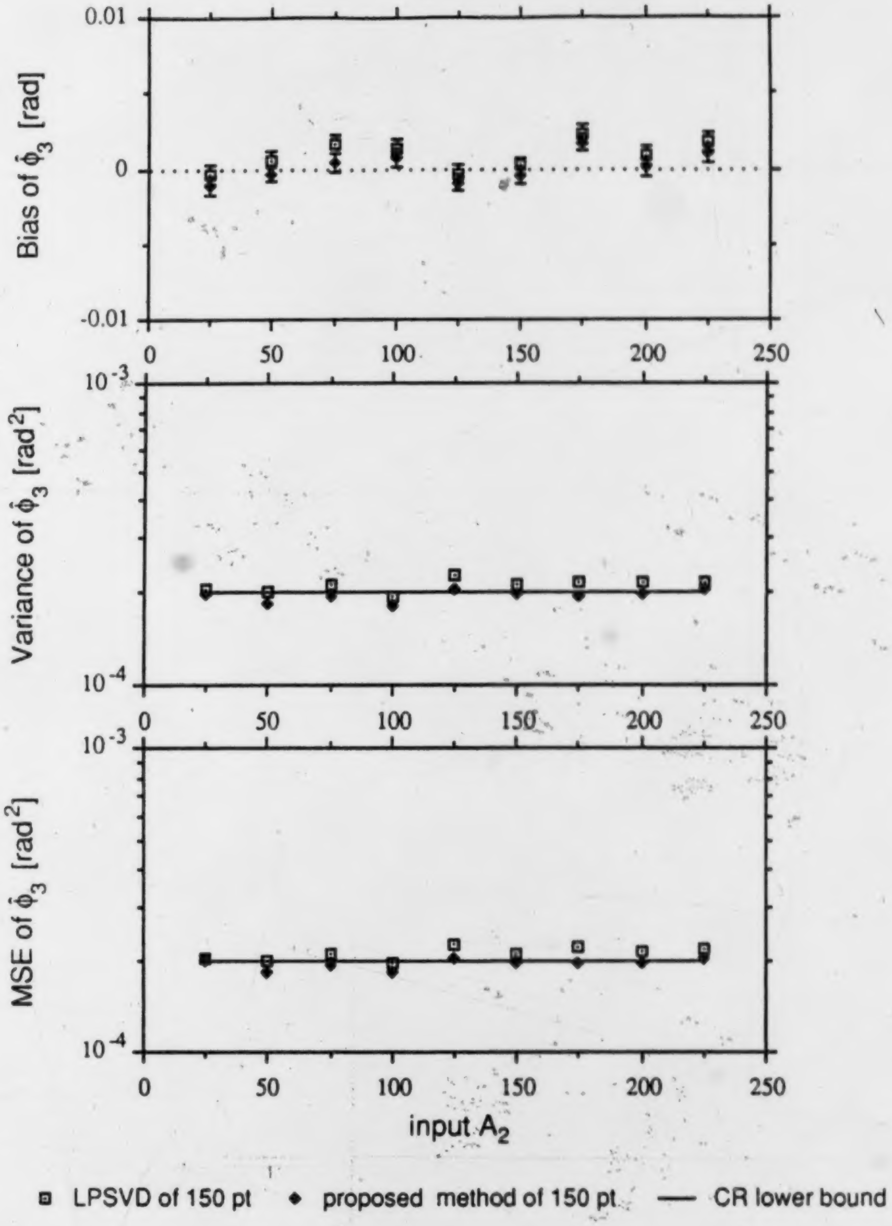


***Figure 5.8.1.11** Phase of the 3-rd peak versus input A_2. For other information, see Figure 5.8.1.4.*

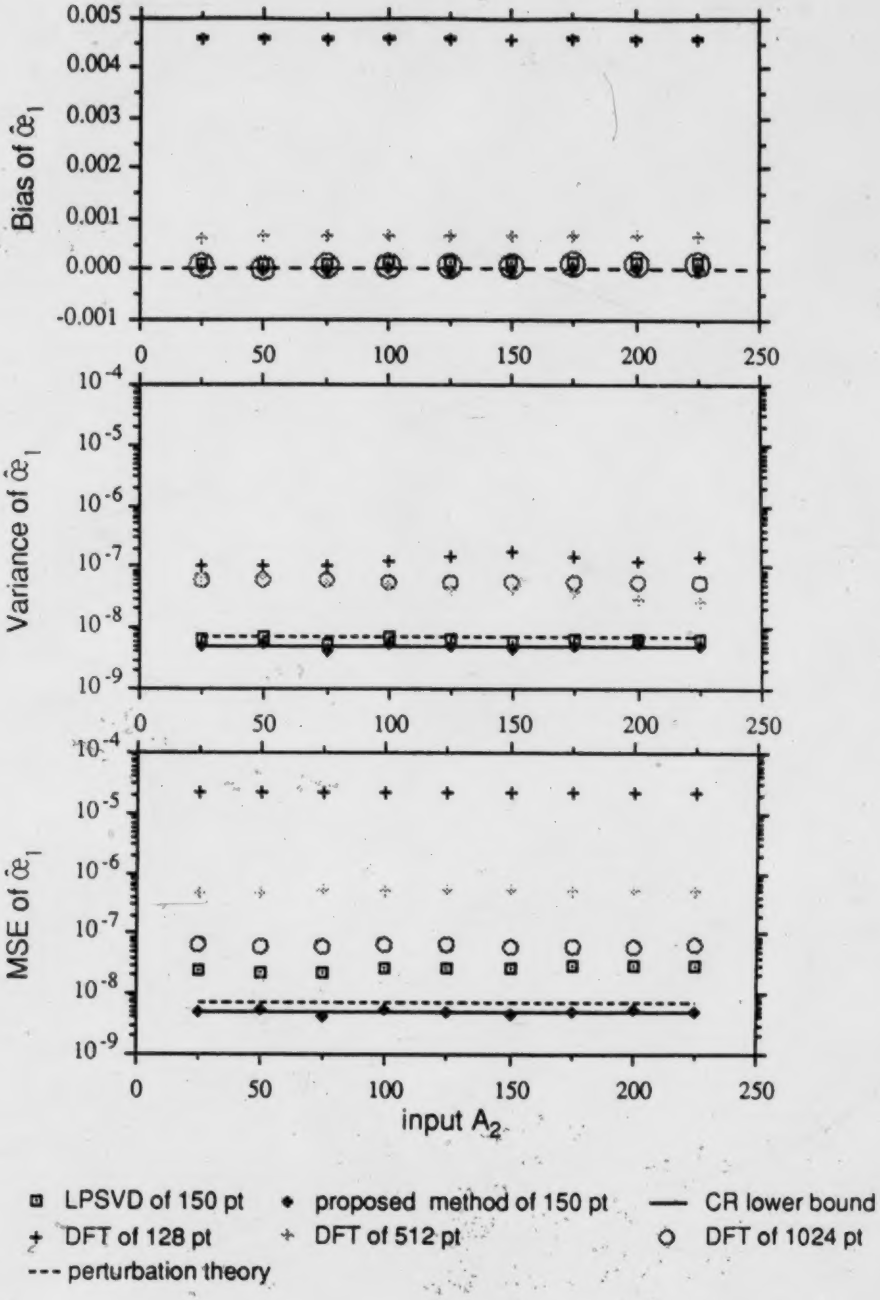


Figure 5.8.1.12 *Normalized damping rate of the 1-st peak versus input A_2. For other information, see Figure 5.8.1.4.*

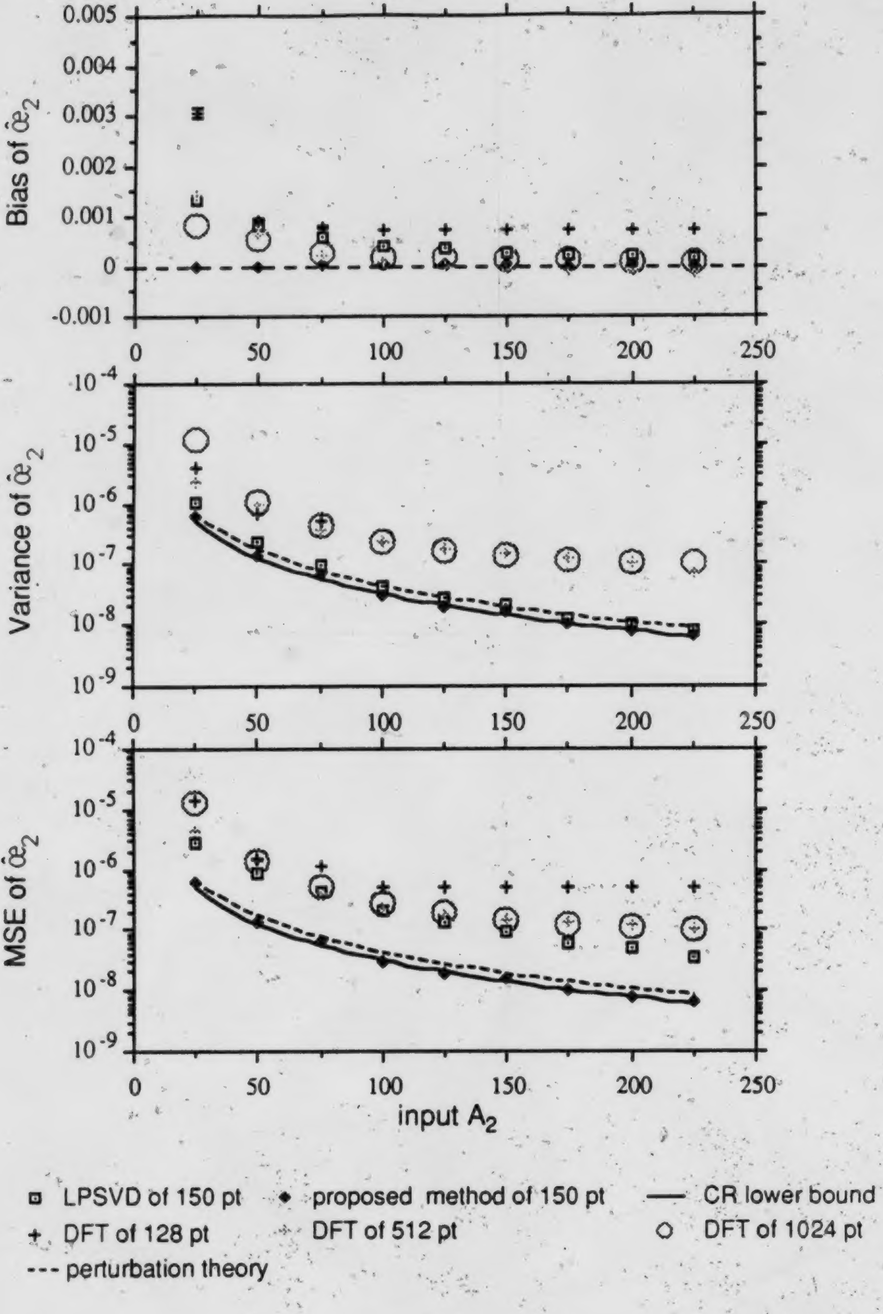


Figure 5.8.1.13 *Normalized damping rate of the 2-nd peak versus input A_2. For other information, see Figure 5.8.1.4.*

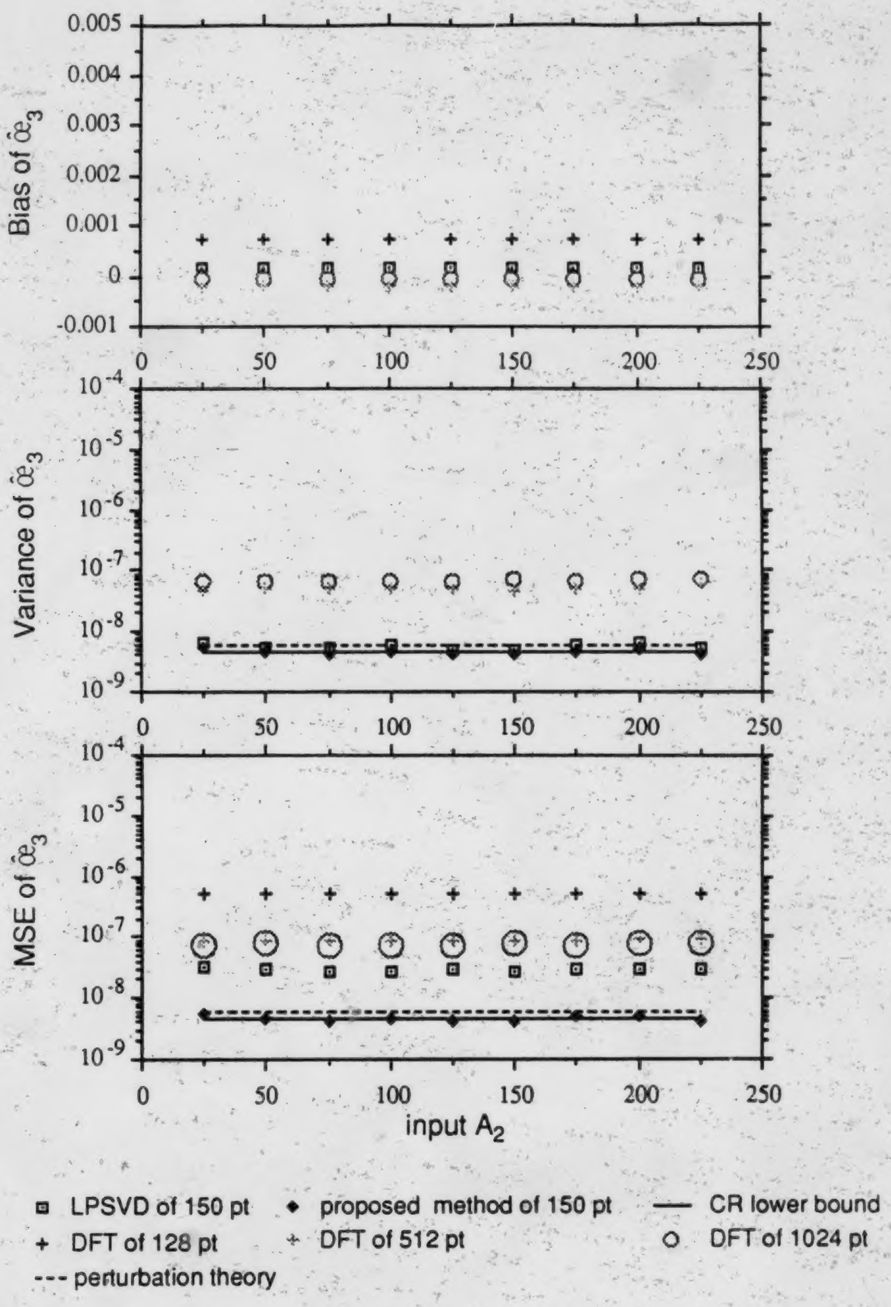


Figure 5.8.1.14 *Normalized damping rate of the 3-rd peak versus input* A_2*. For other information, see Figure 5.8.1.4.*

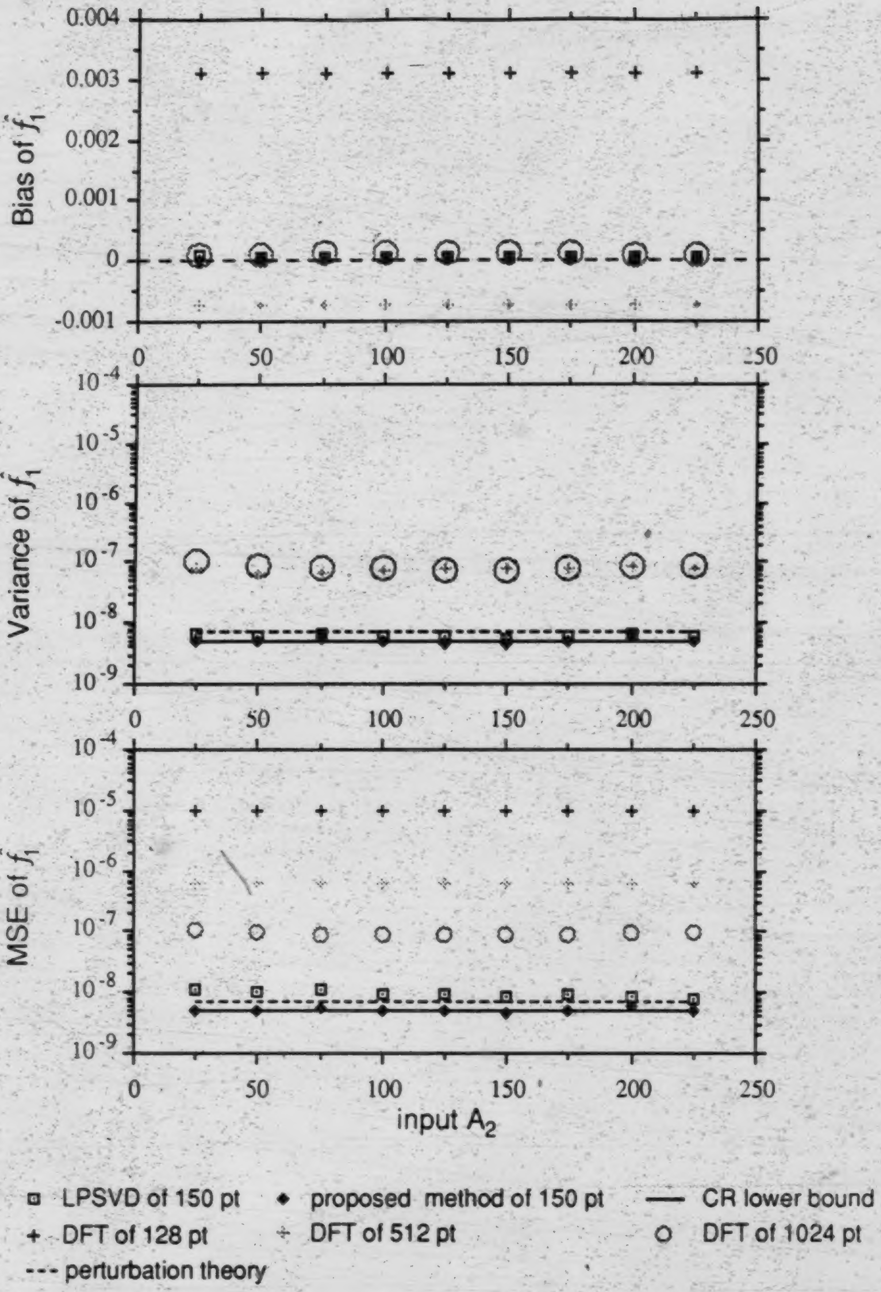


Figure 5.8.1.15 *Normalized frequency of the 1-st peak versus input A_2. For other information, see Figure 5.8.1.4.*

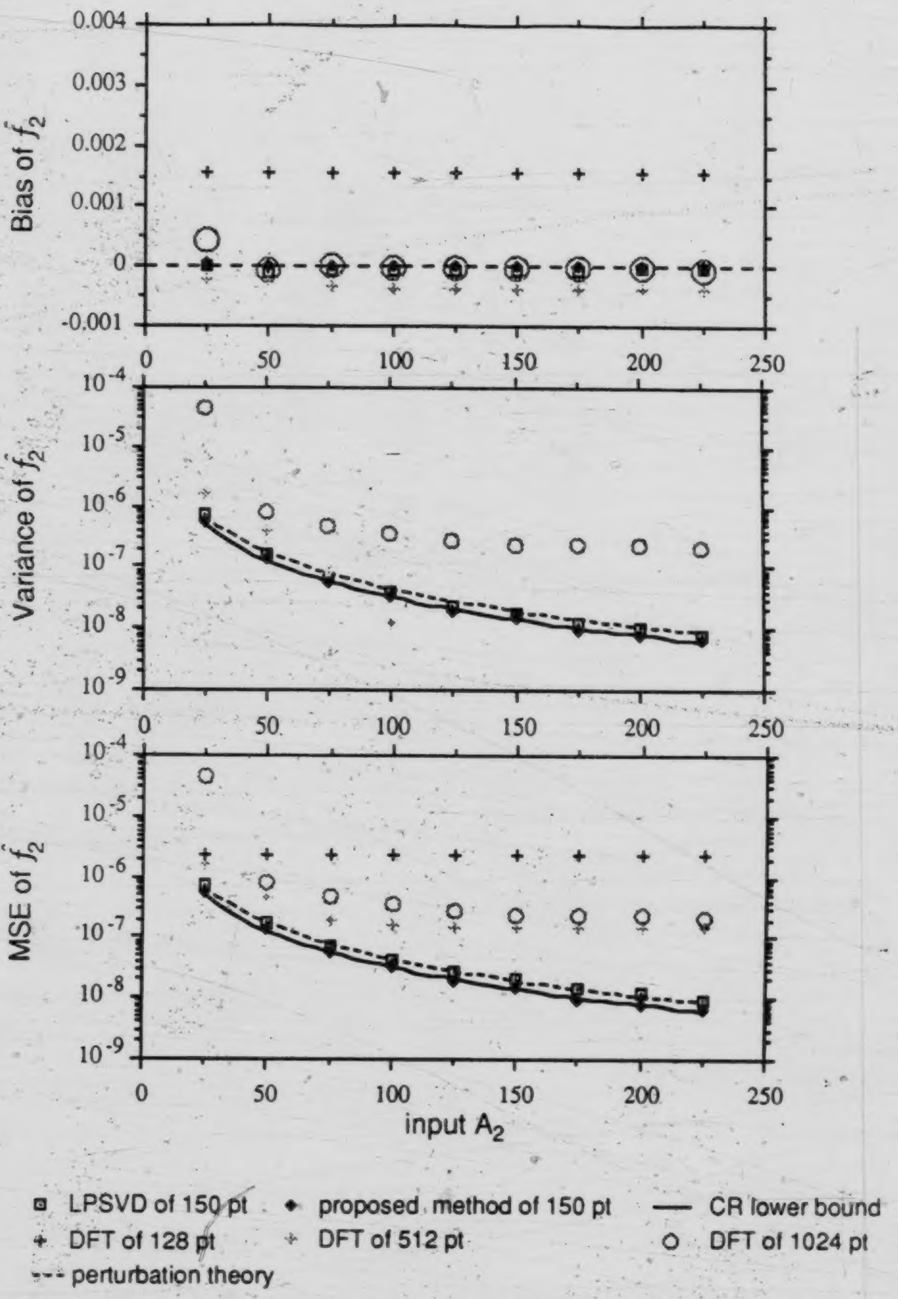


Figure 5.8.1.16 Normalized frequency of the 2-nd peak versus input A_2. For other information, see Figure 5.8.1.4.

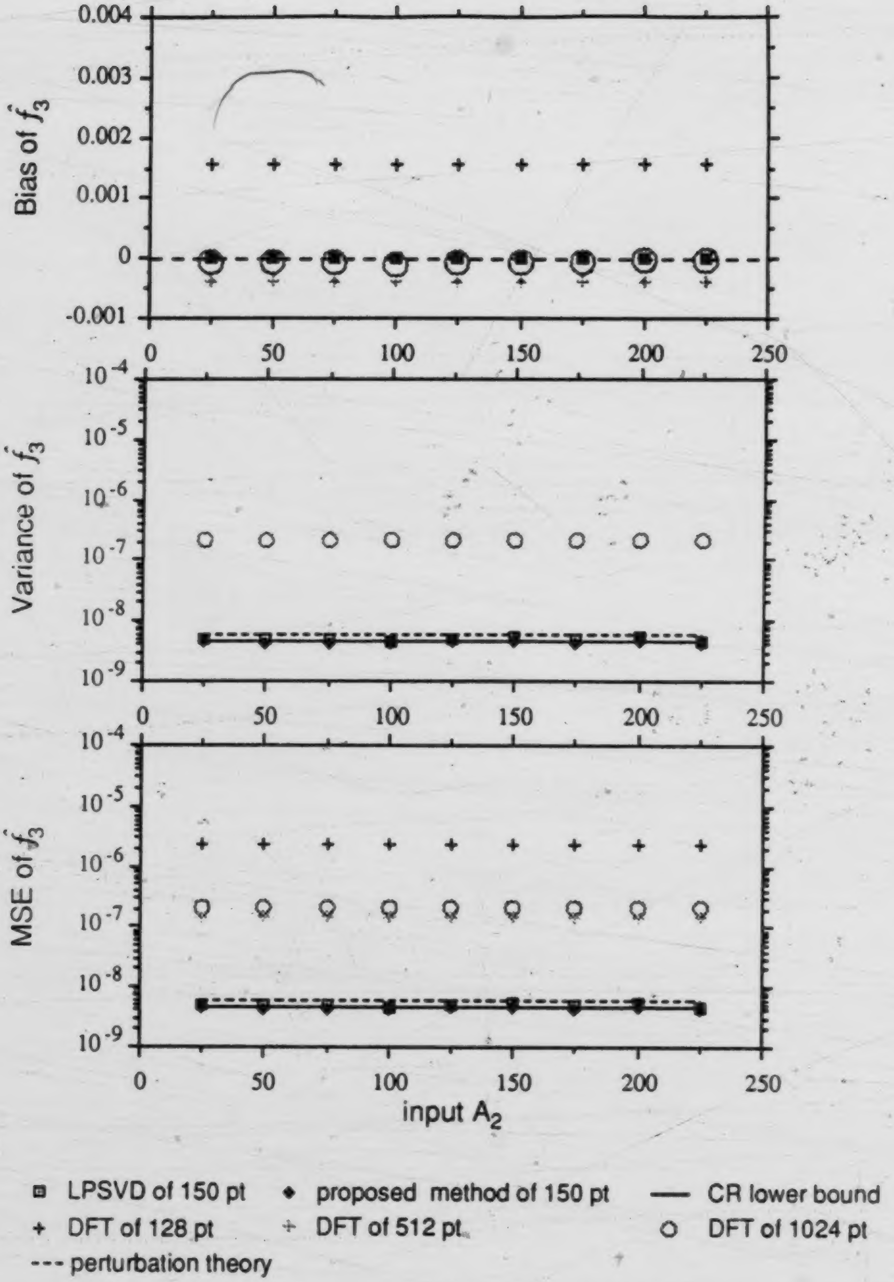


***Figure 5.8.1.17** Normalized frequency of the 3-rd peak versus input A_2. For other information, see Figure 5.8.1.4.*

5.9 Effect of variable baseline (i.e., broad underlying signal)

In many *in-vivo* spectra there is a variable baseline which must be removed before quantitative analysis of the spectral lines can take place. A question is whether this baseline influences the accuracy of the parameter estimates. To address this question simulated FIDs or spectra with increasing baseline components are quantified using the three methods. The kind of baseline we will address is the same as that studied by Nelson and Brown (1989). It comprised a single Lorentzian peak of constant broad linewidth but varying intensity, to which are added two narrow peaks.

Example 5.9.1 :

The time domain parameters: amplitudes A_i, phases ϕ_i, normalized damping rates œ_i and normalized frequencies f_i, of the noise-free signals are as follows:

i	A_i	ϕ_i	œ_i	f_i
1	7	0	0.003183	-0.3
2	5	0	0.003183	-0.1
3	A_3	0	0.127324	0.1

where A_3 varies from 75 to 350. The sampling interval Δt is still 4×10^{-4} sec. Over six hundred independent sets of complex white Gaussian noise, each having independent real and imaginary part of variance $\sigma_w^2 = (1.5)^2 = 2.25$ are added to the noise-free signals for each value of A_3. For the LPSVD method and the proposed method, we used 150 data points. The tentative model order for the LPSVD method was 75. For the DFT method, we use 128, 512 and 1024 data points. Baselines of the DFT spectra are modelled by spline functions.

The range of different spectra considered is shown in Figure 5.9.1.1.

Figure 5.9.1.2 shows the statistical result of the magnitude of the first eight singular values from the LPSVD method. Figure 5.9.1.3 shows the magnitude of the first four singular values versus input time domain amplitude of the third peak, A_3. From the figure, we predict that the LPSVD method will run into trouble when input A_3 is reduced to around 50. In fact, Figures 5.9.1.11 and 5.9.1.12 show that the corresponding frequency estimates already exhibit significant deviation when input A_3 is reduced to 75.

Figures 5.9.1.4 to 5.9.1.12 show the statistical results of the performance measures for the estimates of the spectral parameters from different calculation methods. From the figures, we see that, except for the frequency estimates where input A_3 decreases to 75, the LPSVD method outperforms the 1024-point DFT method. The proposed method, in turn, out performs both the LPSVD method and the DFT method. In addition, the proposed method is able to correct the frequency estimates from the LPSVD method to attain CR lower bounds for input A_3 equal to 75. For input A_3 greater than 75, the variance estimate of the damping rates and frequencies estimated by the LPSVD method are independent of input A_3, which is consistent with property 1 of section 4.2.2. The variance estimate of all parameters estimated by the proposed method attain CR lower bounds for this range of input A_3 being studied. For the DFT method, the amplitude (or area) estimates of the first peak are not affected as much as the second peak which is closer to the high intensity broad peak, by increasing the intensity of the broad peak. A large baseline component caused a reduced time domain amplitude (or frequency domain peak area) estimate, especially for the peak close to the mountain of the baseline peak.

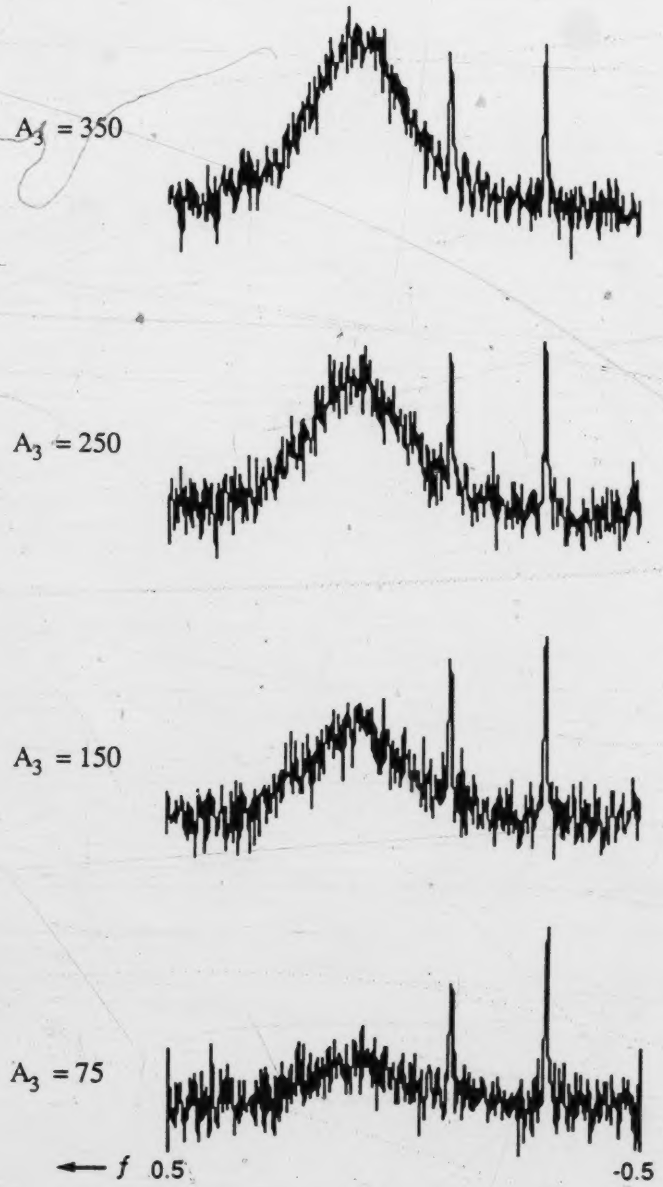


***Figure 5.9.1.1** Some representative DFT spectra for different time domain amplitudes of the 3-rd peak, transformed from 512-point FID data which is composed of three decaying sinusoids plus Gaussian noise. The time domain parameters are: amplitudes $A_1 = 7$, $A_2 = 5$; normalized damping rates $\alpha_1 = \alpha_2 = 0.003183$, $\alpha_3 = 0.127324$; normalized frequencies $f_1 = -0.3$, $f_2 = -0.1$, $f_3 = 0.1$; noise standard deviation 1.5.*

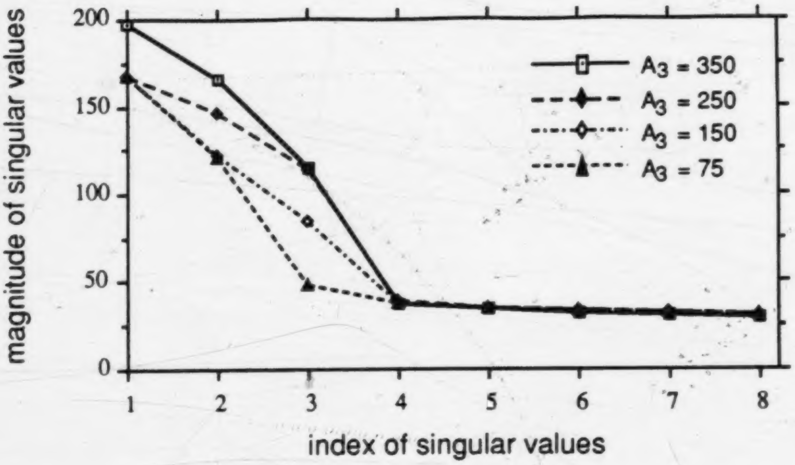


Figure 5.9.1.2 *Statistical result of the first eight singular values of the LPSVD method. A_3 is the input time domain amplitude of the 3-rd peak. Number of data points used is 150. Tentative model order L is 75. For other information, see Figure 5.9.1.1.*

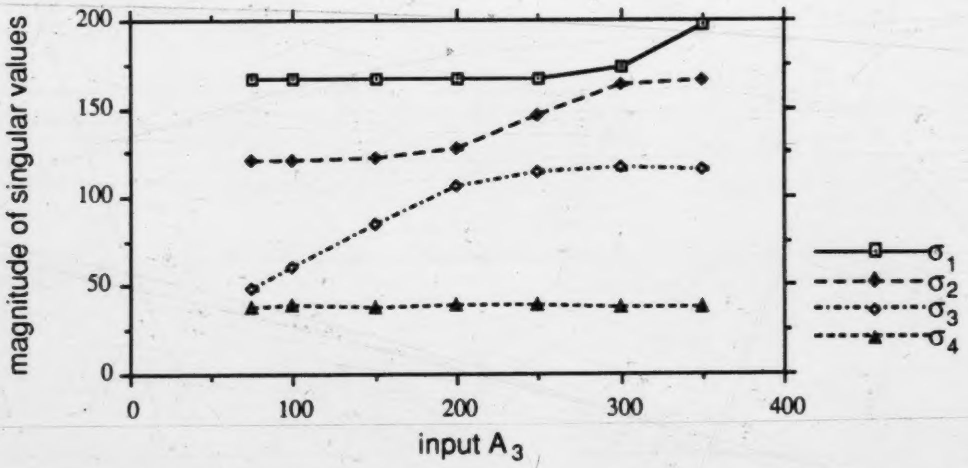


Figure 5.9.1.3 *Magnitude of the first four singular values of the LPSVD method versus input time domain amplitude of the 3-rd peak, A_3. For other information, see Figure 5.9.1.2.*

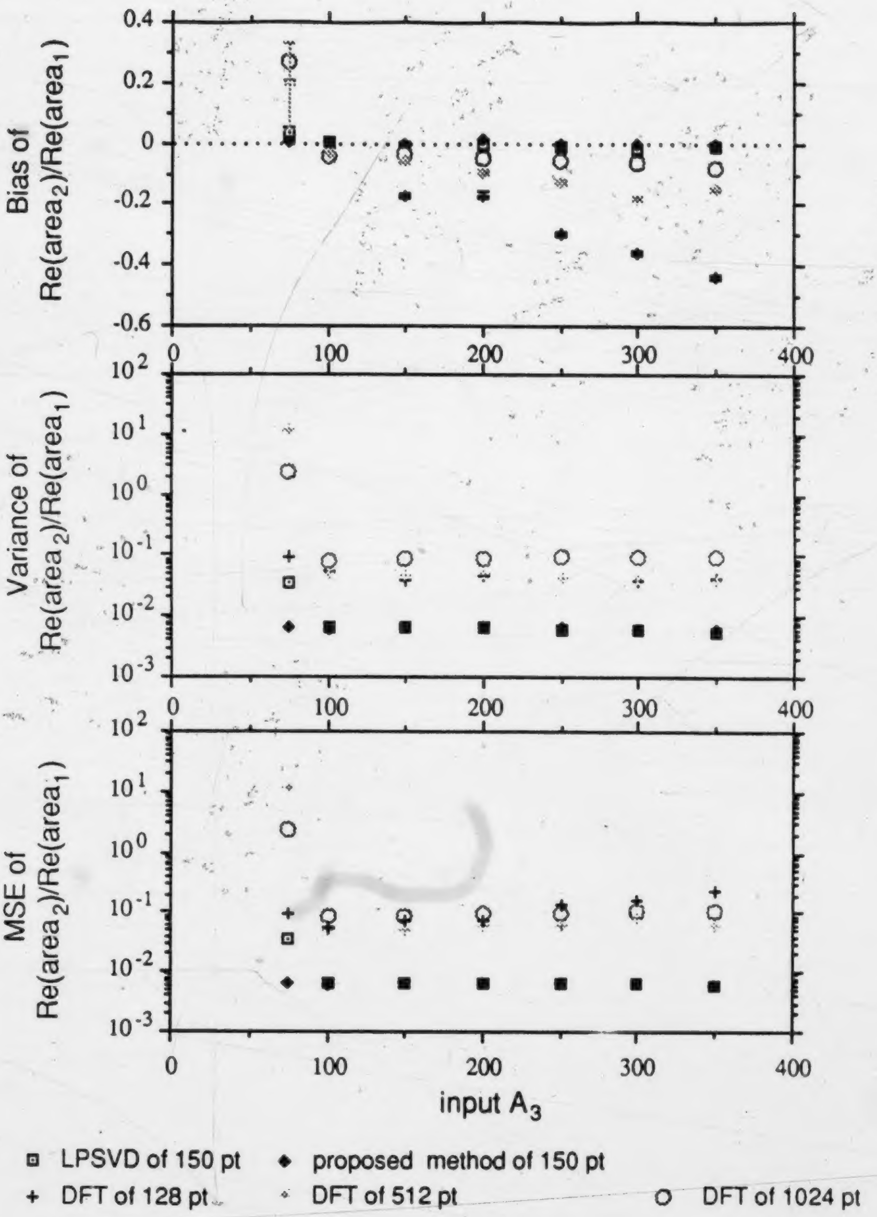


Figure 5.9.1.4 *The area ratio of the 2-nd peak to the 1-st peak from the real spectrum versus input time domain amplitude of the 3-rd peak, A_3. Tentative model order for the LPSVD method is 75. All estimates are obtained from over 600-run simulations. For other information, see Figure 5.9.1.2.*

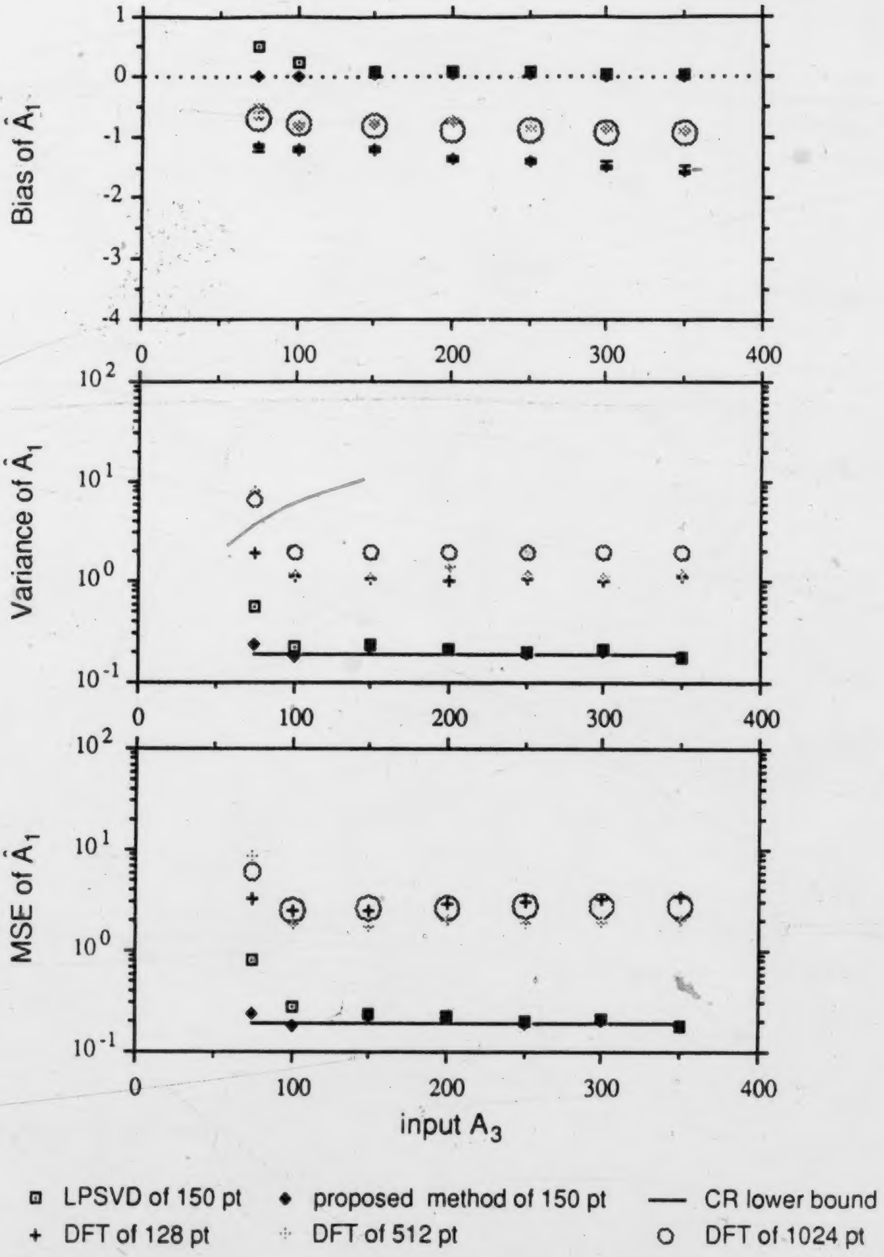


Figure 5.9.1.5 *Time domain amplitude of the 1-st peak versus input A_3. For other information, see Figure 5.9.1.4.*

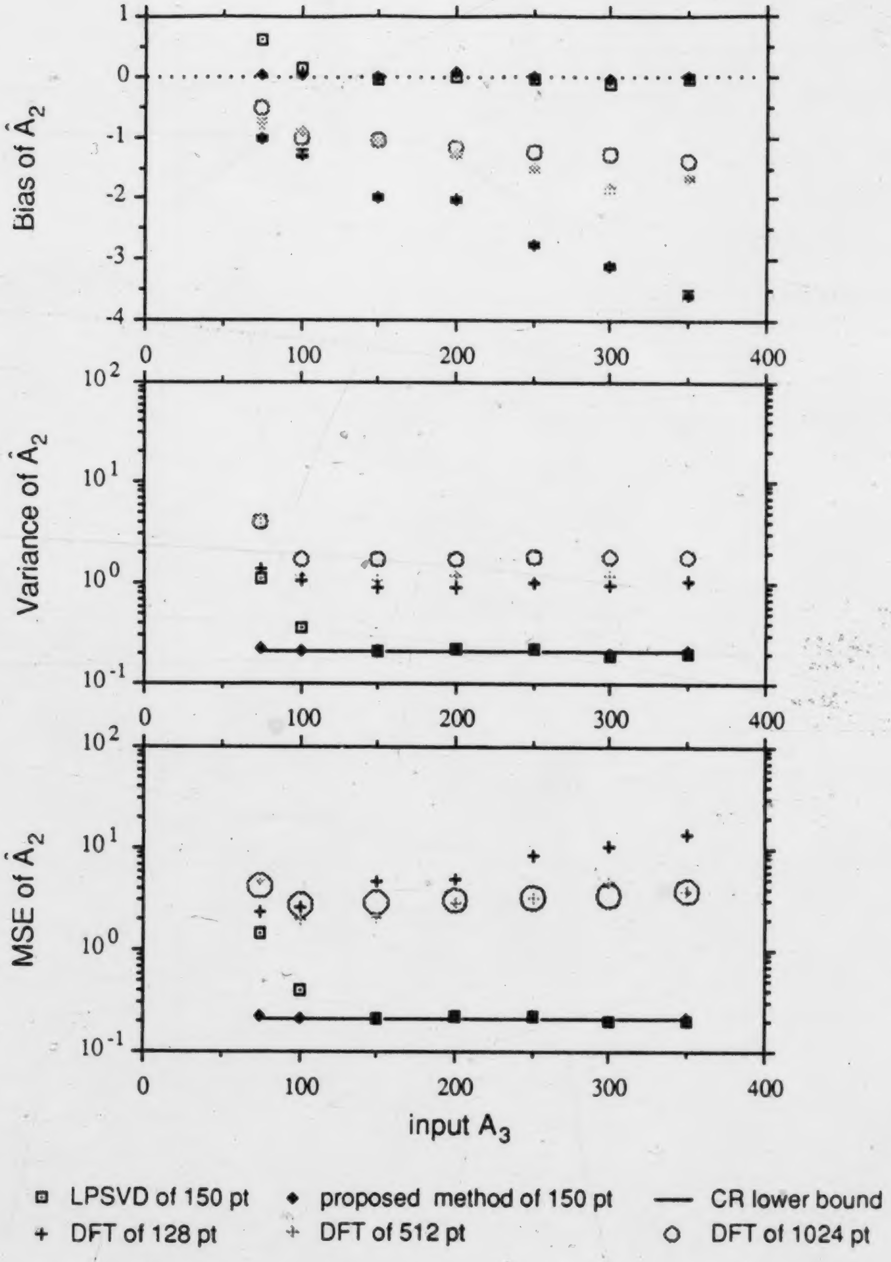


Figure 5.9.1.6 *Time domain amplitude of the 2-nd peak versus input* A_3. *For other information, see Figure 5.9.1.4.*

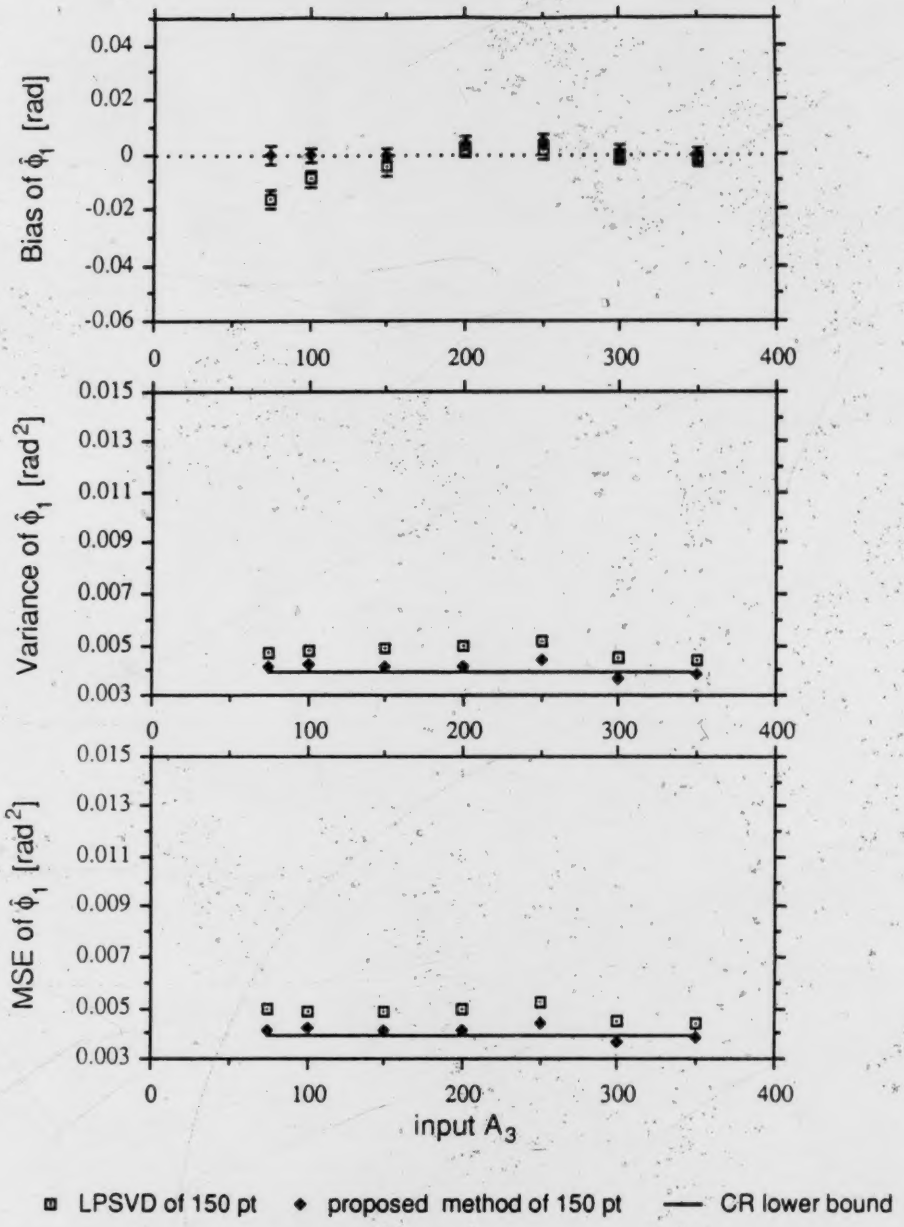


Figure 5.9.1.7 *Phase of the 1-st peak versus input A_3. For other information, see Figure 5.9.1.4.*

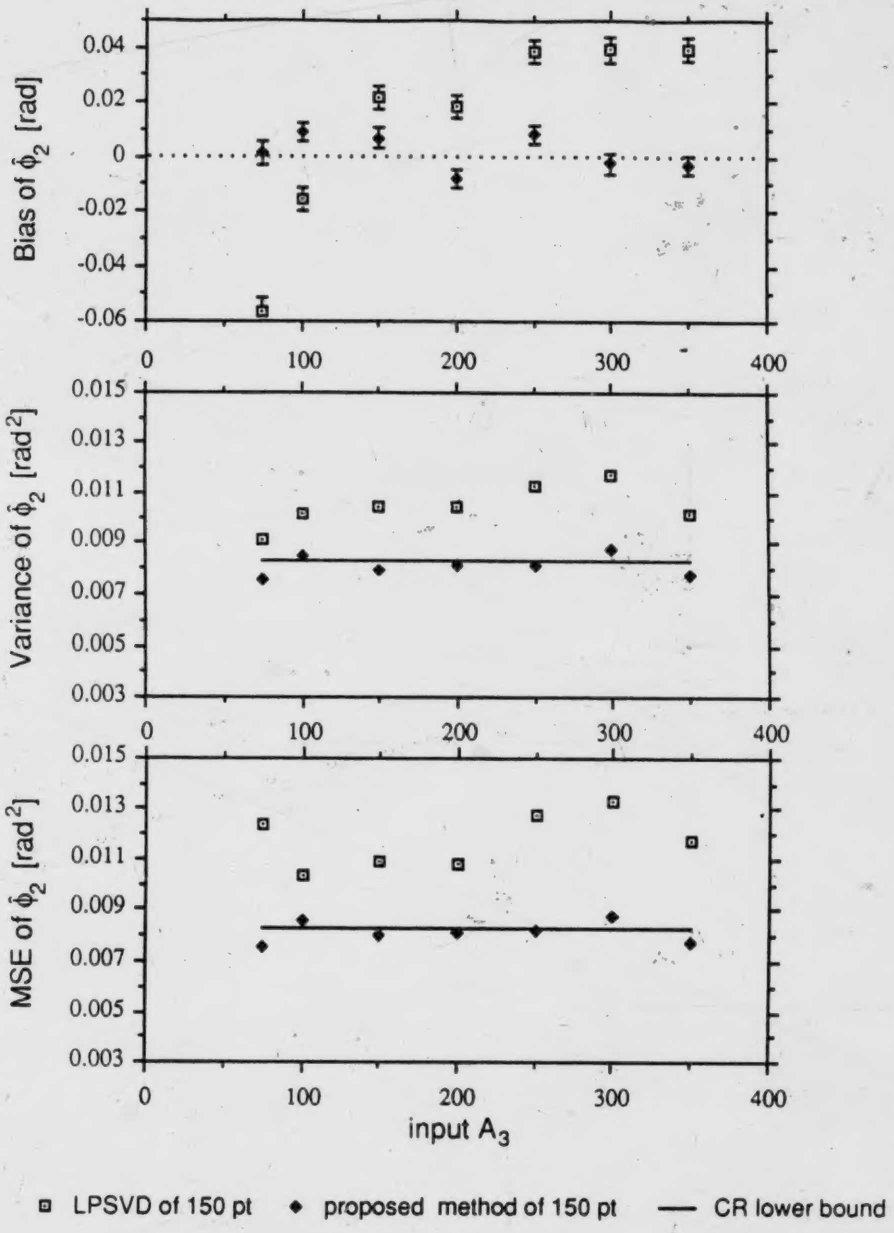


Figure 5.9.1.8 Phase of the 2-nd peak versus input A_3. For other information, see Figure 5.9.1.4.

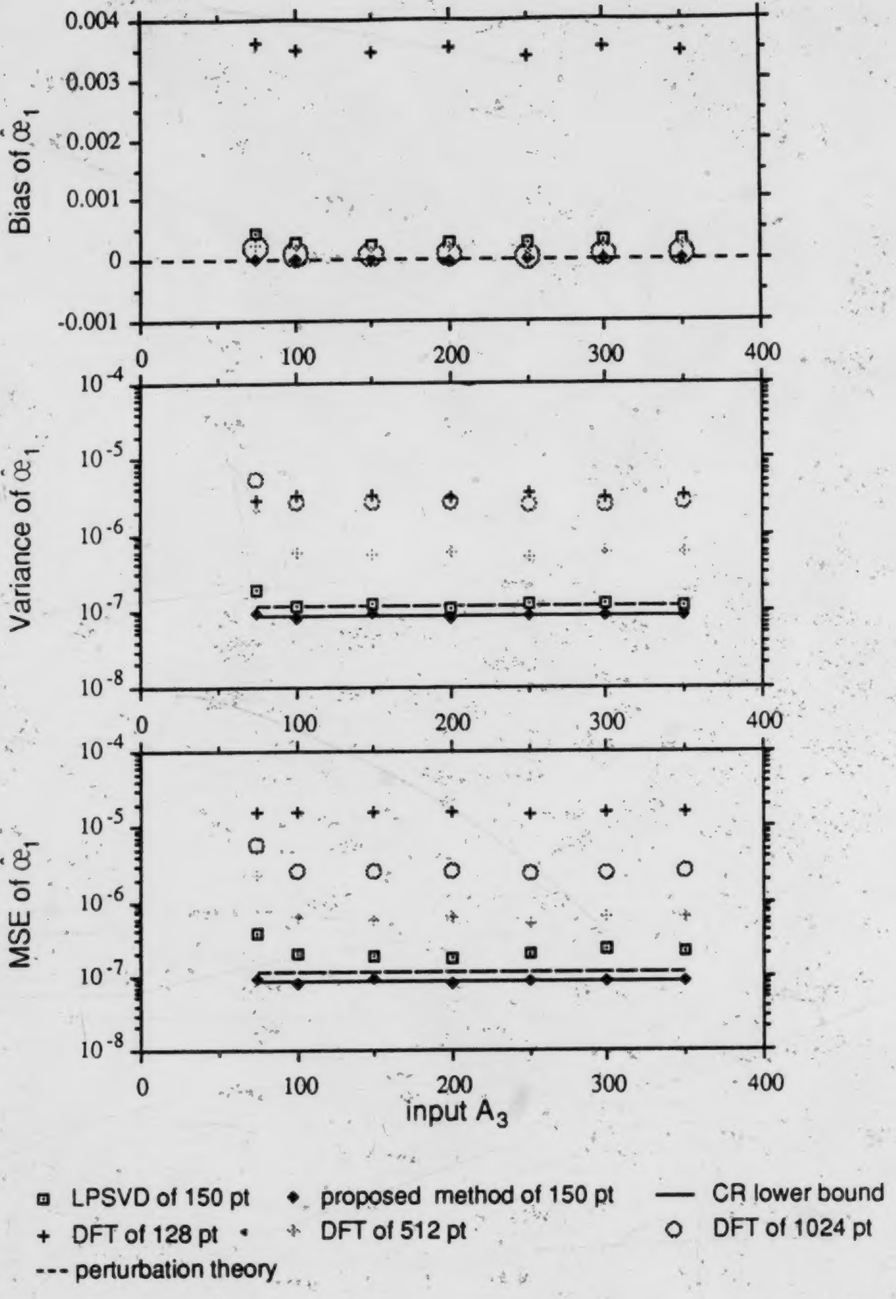


Figure 5.9.1.9 *Normalized damping rate of the 1-st peak versus input A_3. For other information, see Figure 5.9.1.4.*

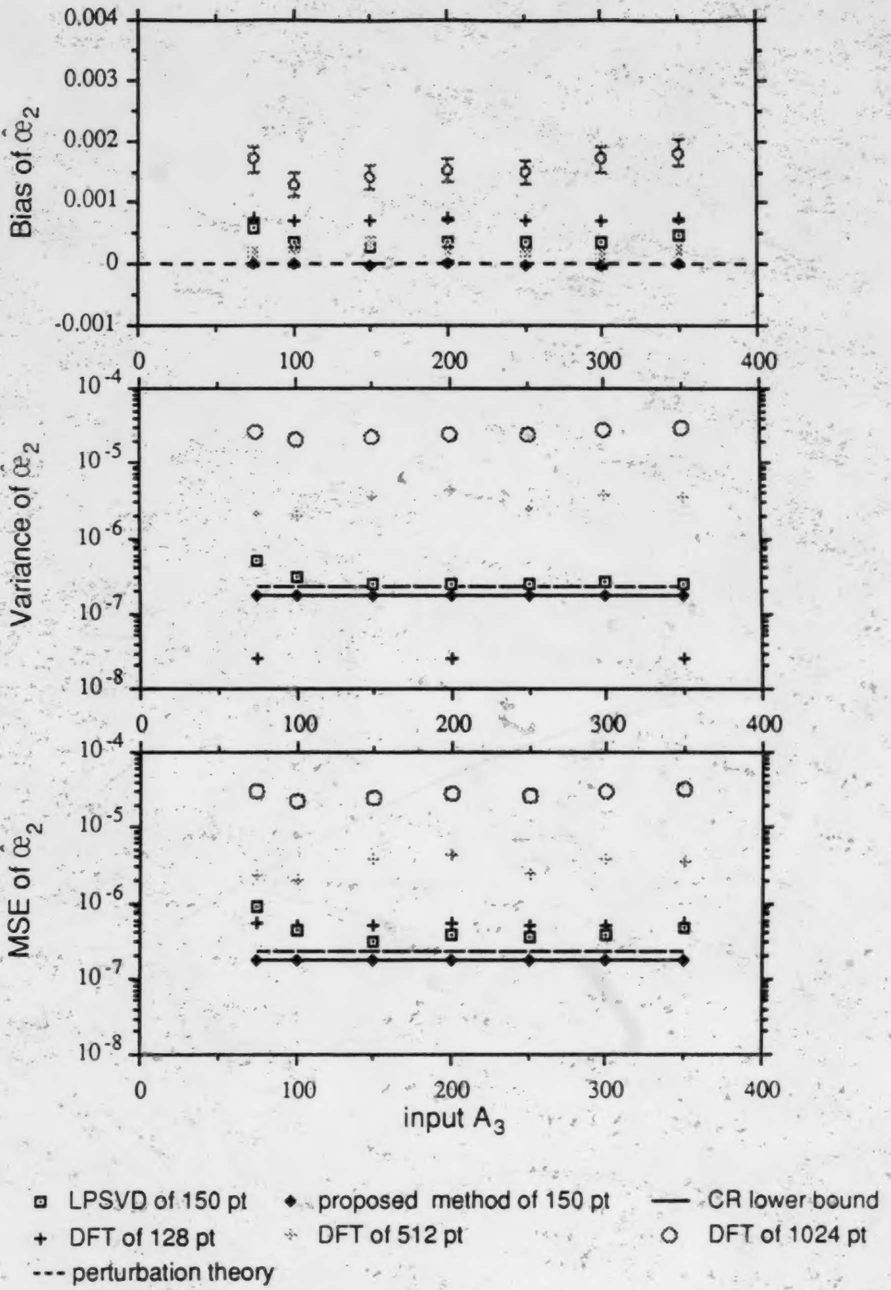


Figure 5.9.1.10 *Normalized damping rate of the 2-nd peak versus input A_3. For other information, see Figure 5.9.1.4.*

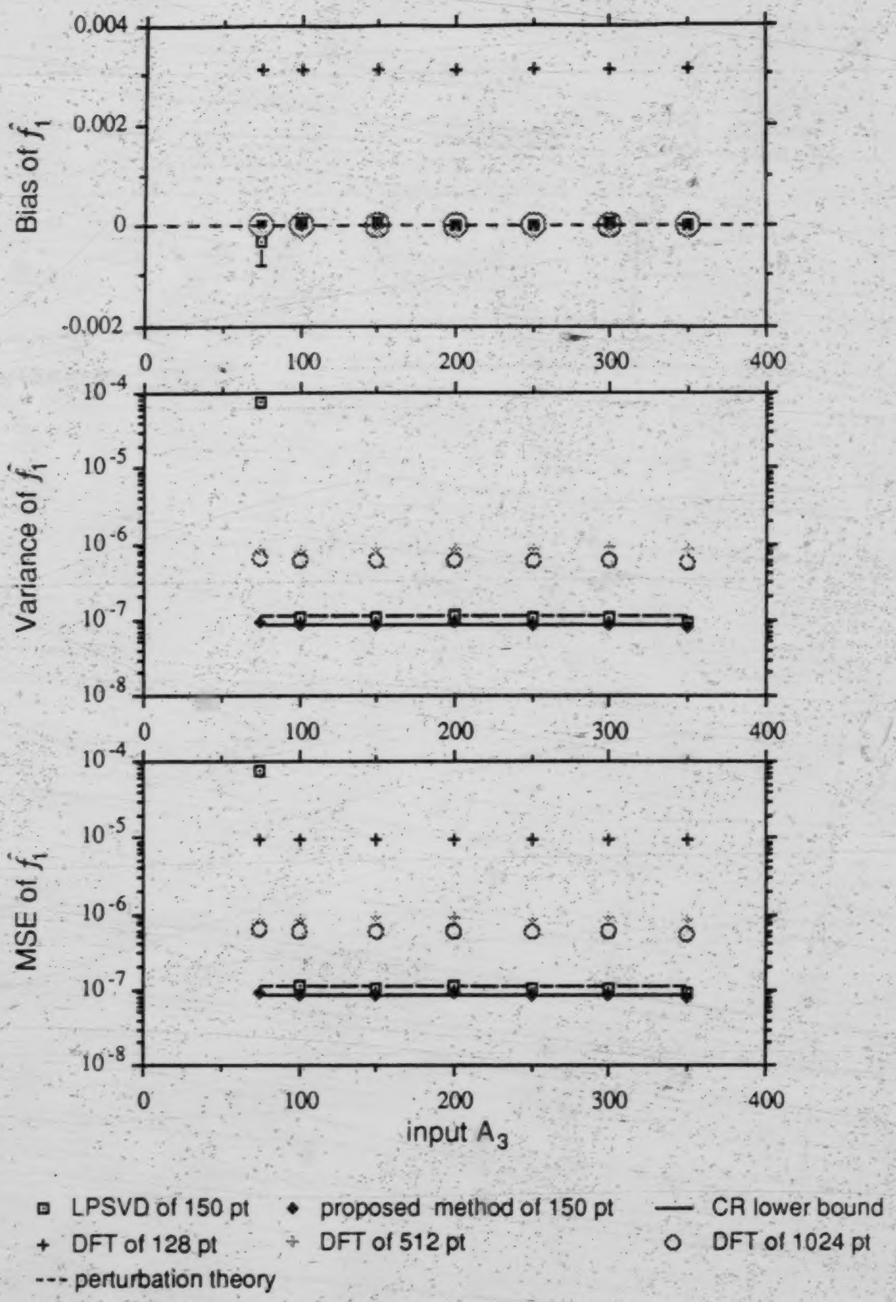


***Figure 5.9.1.11** Normalized frequency of the 1-st peak versus input A_3. For other information, see Figure 5.9.1.4.*

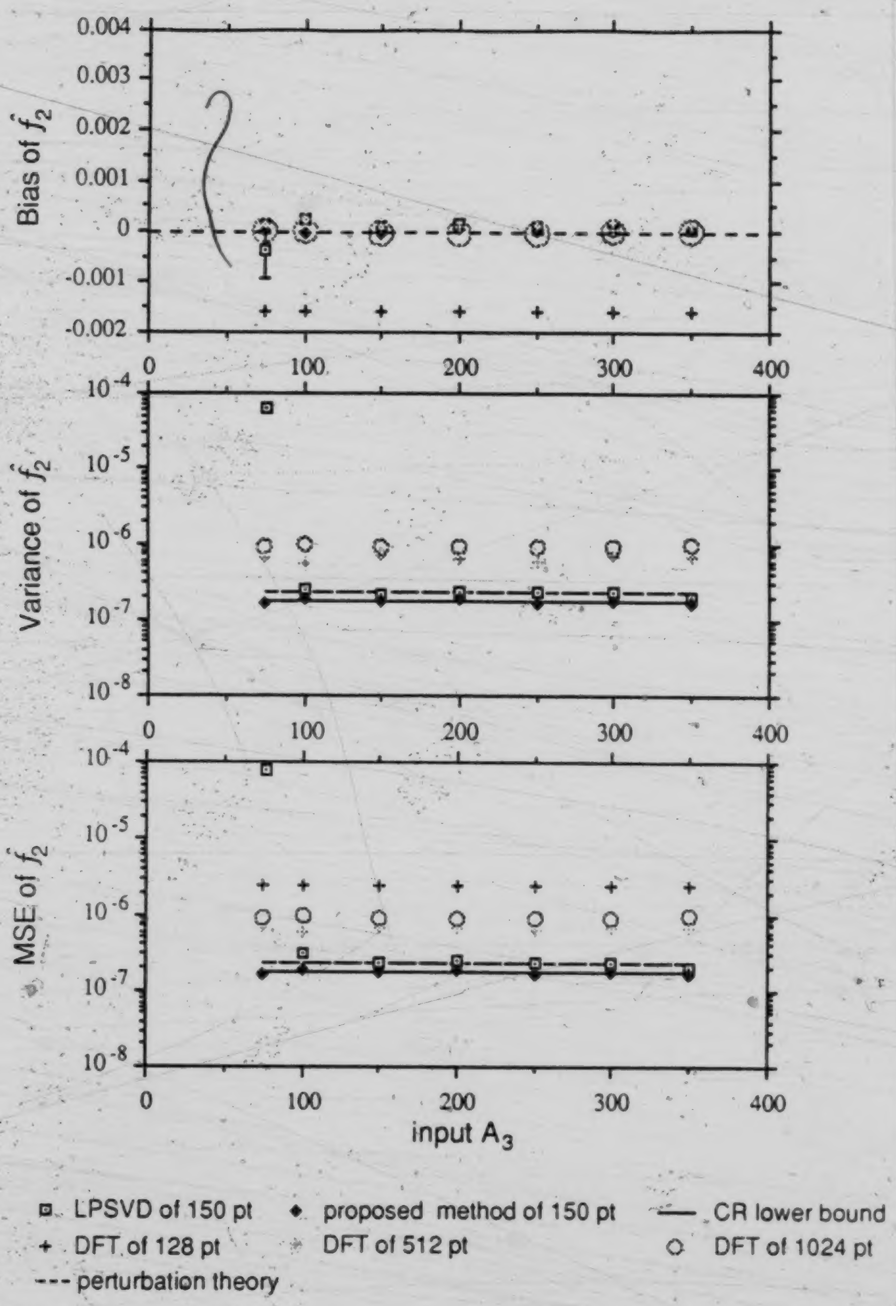


***Figure 5.9.1.12** Normalized frequency of the 2-nd peak versus input A_3. For other information, see Figure 5.9.1.4.*

Chapter 6

Conclusions

Conventional NMR spectroscopy bases its estimations on frequency domain spectra. The transformation from the sampled time domain FID signal to the frequency domain spectrum is done by discrete Fourier transformation (DFT). Poor resolution, implicit windowing and spectral distortion due to the initial delay are some of the disadvantages of this method. Recently, Tufts and Kumaresan developed an advanced time domain method called the TK method or LPSVD method. In section 4.2.2, we have performed first-order perturbation analysis of the LPSVD method. This work was extended from Hua and Sarkar's paper (1988) in which perturbation analysis of the LPSVD method on non-decaying exponential signals was studied. In section 4.3, we developed a proposed method. In Chapter 5, we compared the performance of the DFT, LPSVD and the proposed methods.

The LPSVD method has considerable advantages over the DFT method for the type of spectra considered in Chapter 5, the most important in practical terms being the ability to obtain automatic estimates of the peak parameters. In addition, the method does not assume that the peaks are linearly phase shifted, making it more flexible than the DFT method. Moreover, it is able to resolve components which have overlapping peaks in the frequency domain. In Chapter 5, we discovered that the LPSVD method has another advantage: it is able to estimate the signal parameters more accurately with significantly fewer data points than the DFT method. This observation is particularly valuable in multi-dimensional NMR where accumulation time constraints often lead to limited sampling in at least one of the time dimensions. In a three-dimensional experiment, for example, each complex point in the slowest incrementing time dimension may take as much as an hour to accumulate. Therefore, the LPSVD method can be used to estimate the signal parameters or predict additional points in the time domain data (Gesmar and Led, 1989; Zeng, et al., 1989; Brown and Campbell, 1990; Olejniczak and Eation, 1990). The main disadvantage of the LPSVD method is its specific assumption about lineshape: if the lineshape is not a good model to the data, the results are difficult to interpret.

We discovered that the performance of the LPSVD algorithm can be affected by the following:

(a) The ratio of the tentative model order to the number of data points:

The tentative model order affects the performance of the LPSVD method as demonstrated in Examples 5.2.1 and 5.2.2. If we are given a number of data points N, it may be optimal to choose the tentative model order L to be between 0.4N and 0.8N.

(b) The number of data points:

Keeping the ratio of the tentative model order to the number of data points constant as in Example 5.2.3, we found that the performance of the LPSVD method improves as the number of data points increases, but the improvement seems to "level off". This suggests that it might not be useful to use too many data points for the LPSVD method. The LPSVD method is a very computational intensive algorithm as compare with the DFT method. If we only need a small number of data points to obtain an estimate of the parameters as accurate as, say, the 1024-point DFT method, then this property can save computation time. Of course, in reality, no one knows how many data points are needed to obtain the optimal solution. One way to deal with this problem is as follows. Since the proposed method, which takes the solution from the LPSVD method as an initial estimate for a nonlinear curve fitting routine, is able to give a solution that attains the theoretical lowest CR bound and is unbiased, then we can first use a reasonable number of data points, say 150, to obtain a solution from the proposed method, then use this solution to obtain a plot of the CR lowest bound versus number of data points. If the chosen number of data points does not seem to be optimal, we can rerun the proposed method using a larger number of data points.

(c) The amount of noise:

As shown in Example 5.4.1, noise makes the LPSVD algorithm give biased estimates of the parameters. The biases of the estimated parameters do not seem to be linear with respect to the input time domain noise. As noise increases, the biases become more serious. The variance estimate of the parameters also increase as noise increases. Although, as shown in Example 5.4.1, the biases of some parameters estimated by the LPSVD method are larger than those of the DFT method with 512 or 1024 points at low signal-to-noise, the variance of the parameters estimated by the LPSVD method are about 10 times smaller than those of the DFT method.

(d) Peak separation:

As shown in Example 5.5.1, when the frequency separation between two peaks decreases, the bias estimates of the parameters increase slowly, and the variance estimates of the parameters also increase. The trend of increase in the variance is very much like the trend of the CR bounds. As the normalized frequency separation decreases to about 0.005, the biases of the parameters seem to be more noticeable. However, the LPSVD method still has a higher resolving power than the 1024-point DFT method, which cannot resolve peaks separated by less than 0.006 of a normalized unit. Notice that the resolving power of the DFT method is not only a function of frequency separation and the number of data points in the time domain, but also a function of the linewidths. In Example 5.6.1, we had two peaks separated by a normalized frequency of 0.01. The 1024-point DFT method could not resolve the two peaks as the linewidth of one of the peaks increases to 0.01 of a normalized unit (linewidth = 2 x damping rate), but the LPSVD method was still able to resolve the peaks until the linewidth increased to about 0.032 of a normalized unit. In conclusion, the LPSVD method of 150 points had a higher resolving power than the 1024-point DFT method in these two examples.

(e) Linewidth of the peaks:

As shown in Example 5.6.1, changes in a peak's damping rate affect the performance measures of the parameters of not only that particular peak, but also one close to it. However, the later is affected much less seriously than the former. In Example 5.6.1, we called the peak whose damping rate was changing the second peak. Except for the normalized damping rate and frequency of this changing peak, the variance estimates of the other parameters (including this peak and the other peak) from the LPSVD method increased and then gradually decreased as the input damping rate increased. For the normalized damping rate and normalized frequency of the second peak, the variances continued to increase as their input linewidth increased. The variance estimates of the amplitudes were very similar to those of the corresponding phase. The variance estimate of the normalized damping rate, in turn, was very similar to that of the corresponding normalized frequency.

As mentioned before, the 128-point DFT method could not resolve the two peaks in these simulations. The 512-point and 1024-point DFT method resolved the peaks if the input normalized damping rate of the variable peak was less than 0.0064, while the LPSVD method of only 150 points resolved the two peaks for an input normalized damping rate increase of up to 0.016, 2.5 times wider than the 1024-point DFT could resolve.

(f) Amplitudes of the signal components:

As shown in Example 5.7.1, the LPSVD method is a biased estimator. The bias estimates of the phase and normalized damping rate of a peak (we called this the second peak) increased as the input time domain amplitude A_2 decreased. The biases of the other parameters did not change significantly as A_2 changed. The variances of the normalized damping rate and frequency of the constant peak (we called this the first peak) were not affected by changing A_2, but those of the second peak were inversely proportional to $|A_2|^2$. This observation was consistent with the perturbation analysis given in section 4.2.2. In general, the estimate performance of the first peak was not affected by changing the input A_2.

Compared to the 1024-point DFT method, variances of the parameters obtained from the LPSVD algorithm of 150 points were significantly smaller. The biases of the parameters from the two methods were very close to one another. But, if we used only 512 points, the DFT method behaved poorer than the LPSVD method of 150 points.

Both methods, together with the proposed method, deteriorated significantly for input A_2 less than 25 in this simulation.

(g) Broad underlying peak:

From Example 5.9.1, we saw that except when the broad underlying signal was small such that the frequency estimates of the peaks had larger biases and variances than the estimates obtained from the 1024-point DFT method, the LPSVD method out performed the 1024-point DFT method. When the broad underlying signal became larger, the variance estimate of the damping rates and frequencies from the LPSVD method were not affected much by the intensity of the broad peak. This is consistent with property 1 of section 4.2.2.

Since the LPSVD method produces biased estimates in general, we developed a method to reduce these biases. We called it the proposed method. It uses estimates from the LPSVD algorithm as initial guesses. It turns out that the proposed method improves the estimates (including biases and variances) of the LPSVD method. The improvement is due to the fact that maximum likelihood estimates of the parameters of M exponentially decaying sinusoidal signals embedded in independent, zero mean, complex valued, Gaussian random noise is equivalent to obtaining the parameters which minimize the least-squares error. It is difficult to use coarse searches to obtain a global minimum least-squares error, so we suggested using the results from the LPSVD algorithm as initial estimates. From the simulation studies in Chapter 5, the results from the proposed method, in general, are unbiased and attain the lowest theoretical Cramer-Rao bounds. Its solutions are, in general, better than those of the LPSVD method, which in turn, are better than those of the 1024-point DFT method.

Although many questions have found their answers in this thesis, new problems are always emerging faster than solutions. A few examples of questions that might be worthy of more study are the following:

(a) Almost all examples in Chapter 5 showed that the LPSVD method is biased. Since our first-order perturbation analysis suggests that it is not biased, higher-order perturbation theory is needed for analytical comparison with the simulation results.

(b) Although both the LPSVD and the proposed methods have better performance than the 1024-point DFT method in general, they fail in situations when the signal and noise singular values are not well separated. Research to extend the operational range further is needed. Application of robust estimation techniques may extend the operational range further.

(c) The LPSVD algorithm is very computational intensive compare to the DFT method. It is about 30 times slower, so improving the speed of the algorithm is needed for real-time data analysis.

(d) The proposed method minimizes the residual error square; other error equations could also be minimized.

(e) Although linear-prediction methods have recently been developed in multi-dimensional NMR spectroscopy, their behaviors have not been studied extensively. In fact, as far as we know, perturbation analysis has not been developed.

References

Abragam, A., *Principles of Nuclear Magnetism*, Clarendon Press, 1961.

Albert, A., *Regression and the Moore-Penrose Pseudo-inverse*, Mathematics in Science and Engineering: vol. 94, edited by Richard Bellman, Academic Press, 1972.

Barford, N.C., *Experimental measurements : precision, error and truth*, John Wiley & Sons, second edition, 1985.

Barlett, M.S., "Some examples of statistical methods of research in agriculture and applied biology," *Jour. Roy. Stat. Soc (Suppl.)*, vol. 4, 1937, p. 137.

Barkhuijsen, H., de Beer, R., Bovee, W.M.M.J., van Ormondt, D., "Retrieval of Frequencies, Amplitudes, Damping Factors and Phases from Time-Domain Signals Using a Linear Least-Squares Procedure," *Journal of Magnetic Resonance*, vol. 61, 1985, pp. 465-481.

Bjerhammar, A., *Theory of errors and generalized matrix inverses*, Elsevier Scientific Publishing Company, Amsterdam, vol. 12, 1973.

Boullion, T.L. and Odell, P.L., *Proceedings of the Symposium on Theory and Application of Generalized Inverses of Matrices*, held at the Department of Mathematics, Texas Technological College, Lubbock, Texas, March 1968, Texas Technological College Mathematics Series, No. 4, Texas Tech. Press, Lubbock.

Bovey, F.A., *Nuclear Magnetic Resonance Spectroscopy*, Academic Press, 1988.

Bresler, Y. and Macovski, A., "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-34, No. 5, October 1986, pp. 1081-1089.

Brown, D.E. and Campbell, T.W., "Enhancement of 2D NMR Spectra Using Singular Value Decomposition," *Journal of Magnetic Resonance*, vol. 89, 1990, pp. 255-264.

Cohen-Tannoudji, Diu, B. and Laloe, F., *Quantum Mechanics*, Vol. I, John Wiley & Sons, Hermann, 1977.

Emsley, J.W., Feeney, J. and Sutcliffe, L.H., *High Resolution Nuclear Resonance Spectroscopy*, vol 1, Pergamon Press, 1965.

Ernst, R.R., Bodenhausen, G. and Wokaun, A., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford Science Publication, Oxford, 1988.

Fante, R., *Signal Analysis and Estimation : An Introduction*, John Wiley & Sons, 1988.

Farrar, T.C., Becker, E.D., *Pulse and Fourier Transform NMR : Introduction to Theory and Methods*, Academic Press, 1971.

Frame, J.S., "Matrix functions and applications. I. Matrix operations and generalized inverses," *IEEE Spectrum*, vol. 1, 1964, pp. 209-220.

Gadian, D.G., *Nuclear magnetic resonance and its applications to living systems*, Clarendon Press, 1984.

Gesmar, H. and Led, J.J., "Two-Dimensional Linear-Prediction NMR Spectroscopy," *Journal of Magnetic Resonance*, vol. 83, 1989, pp. 53-64.

Greville, T.N.E., "The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations," *SIAM Review*, vol. 1, 1959, pp. 38-43.

Golub, G.H. and Van Loan, C.F., *Matrix Computations*, The Johns Hopkins University Press, 1985.

Harris, R.K., *Nuclear Magnetic Resonance Spectroscopy: A Physiochemical View*, Longman Scientific & Technical, 1986.

Hua, Y. and Sarkar, T., "Perturbation Analysis of TK Method for Harmonic Retrieval Problems," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-36, No. 2, February 1988, pp. 228-240.

Jackson, L.B., *Digital Filters and Signal Processing*, second edition, Klumer Academic Publishers, 1989.

Jardetzky, O. and Roberts, G.C.K., *NMR in Molecular Biology*, Academic Press, Florida, 1981.

Klema, V.C. and Laub A.J., "The Singular Value Decomposition : Its Computation and some Applications," *IEEE Trans. Automatic Control*, vol. AC-25, No. 2, April 1980, pp. 164-176.

Kumaresan, R., "Estimating the Parameters of Exponentially Damped or Undamped Sinusoidal Signals in Noise," Ph.D. dissertation, University of Rhode Island, 1982.

Kumaresan, R., "On the zeros of the Linear Prediction-Error Filter for Deterministic Signals," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-31, No. 1, February 1983, pp. 217-220.

Kumaresan, R., Scharf, L.L. and Shaw, A.K., "An Algorithm for Pole-Zero Modeling and Spectral Analysis," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-34, No. 3, June 1986, pp. 637-640.

Kumaresan, R. and Shaw, A.K., "High Resolution Bearing Estimation without Eigen-Decomposition," *Proc. IEEE ICASSP 85*, Tampa, FL, 1985, pp. 576-579.

Kumaresan, R. and Tufts, D.W., "Estimating the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-30, No. 6, December 1982, pp. 833-840.

Lang, S. and McClellan, J., "A Simple Proof of Stability for All-Pole Linear Prediction Models," *Proc. IEEE*, vol. 67, No. 5, May 1979, pp. 860-861.

Makhoul, J., "Linear Prediction : A Tutorial Review," *Proc. IEEE*, vol. 63, No. 4, April 1975, pp. 561-580.

Marcus, M. and Minc, H., *A Survey of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, Boston, 1964.

Marquardt, D.W., "An algorithm for least squares estimation of nonlinear parameters," *SIAM J.*, vol 11, 1963, pp. 431-441,.

Moore, E.H., Abstract. *Bull. Amer. Math. Soc.* vol. 26 (1920), pp. 394-395.

Nashed, M.Z., *Generalized Inverses and Applications*, Proceedings of an Advanced Seminar sponsored by the Mathematics Research Center, the University of Wisconsin-Madison, October 8-10, 1973, Academic Press, 1976.

Nelson, S.J. and Brown, T.R., "The Accuracy of Quantification from 1D NMR Spectra Using the PIQABLE Algorithm," *Journal of Magnetic Resonance*, vol. 84, 1989, pp. 95-109.

Olejniczak, E.T. and Eation, H.L., "Extrapolation of Time-Domain Data with Linear Prediction Increases Resolution and Sensitivity," *Journal of Magnetic Resonance*, vol. 87, 1990, pp. 628-632.

Oppenheim, A.V. and Schafer, R.W., *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.

Pakula, L. and Kay, S., "Simple Proofs of the Minimum Phase Property of the Prediction Error Filter," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-31, No. 2, April 1983, p. 501.

Parthasarathy, S. and Tufts, D.W., "Maximum Likelihood Estimation of Parameters of Exponentially Damped Sinusoids," *Proceedings of the IEEE*, vol. 73, No. 10, October 1985, pp. 1528-1530.

Pisarenko, V.F., "The Retrieval of Harmonics from a Covariance Function," *Geophys. J. Roy. Astron. Soc.*, vol. 33, 1973, pp. 347-366.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., *Numerical recipes : the art of scientific computing*, Cambridge University Press, 1986.

Pringle, R.M. & Rayner, A.A., *Generalized inverse matrices with Applications to Statistics*, Griffin's Statistical Monographs and Courses, No. 28, edited by Alan Stuart, Charles Griffin & Company Limited, 1971.

Rao, C.R. and Mitra, S.K., *Generalized Inverse of Matrices and its Applications*, John Wiley & Sons, 1971.

Rao, C.R., *Linear Statistical Inference and its applications*, John Wiley & Sons, New York, 1965.

Reddi, S.S., "Multiple Source Location - A Digital Approach," *IEEE Trans. Aerosp. and Electron. Syst.*, vol. AES-15, January 1979, pp. 95-105.

Rife, D.C. and Boorstyn, R.R., "Multiple tone parameter estimation from discrete time observations," *Bell Syst. Tech. J.*, vol. 55, November 1976, pp. 1389-1410.

Rohde, C.A., "Generalized inverses of partitioned matrices," *J. Soc. Indust. Appl. Math.*, vol. 13, 1965, pp. 1033-1035.

Sheffield, R.D., "A general theory for linear systems," *Amer. Math. Monthly*, vol. 65, 1958, pp. 109-111.

Shinnar M. and Eleff, S.M., "A Generalized Pisarenko Technique for Analyzing the NMR FID," *Journal of Magnetic Resonance*, vol. 76, 1988, pp. 200-208.

Slichter, C.P., *Principles of Magnetic Resonance*, Haper and Row, New York, 1963.

Stanley, W.D., Dougherty, G.R. and Dougherty R., *Digital Signal Processing* (2nd edition), Reston Publishing Company, Prentice-Hall, 1984.

Stewart, G.W., *Introduction to Matrix Computations*, Academic Press, Inc., 1973.

Stoica, P. and Nehorai, A., "On Stability and Root Location of Linear Predication Models," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-35, No. 4, April 1987, pp. 582-584.

Thomas, L.J., *Nuclear Magnetic Resonance in Biochemistry: Principles and applications*, Academic Press, 1975.

Tufts, D.W., "Singular Value Decomposition and Improved Frequency Estimation Using Linear Prediction," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. ASSP-30, No. 4, August 1982, pp. 671-675.

Usmani, R.A., *Applied Linear Algebra*, Marcel Dekker, Inc., 1987.

Wedin, P.A., "Perturbation Theory for Pseudo-inverses," *B.I.T.*, pp. 217-232.

Zeng, Y., Tang, J., Bush, C.A. and Norris, J.R., "Enhanced Spectral Resolution in 2D NMR Signal Analysis Using Linear Prediction Extrapolation and Apodization," *Journal of Magnetic Resonance*, vol. 83, 1989, pp. 473-483.

Appendix 1 Theoretical autocorrelation matrix of M complex exponentially damped sinusoids

Assuming the phases of M complex exponentially damped sinusoids are independent random variables uniformly distributed on $[0,2\pi)$, it is shown that the theoretical L x L autocorrelation matrix is given by

$$\begin{aligned} \boldsymbol{R}_{ss} &= \sum_{i=1}^{M} A_i^2 \mathbf{e}_{ci} \mathbf{e}_{ci}^H , \\ &= \boldsymbol{E} \boldsymbol{P} \boldsymbol{E}^H , \end{aligned} \qquad \text{(A1.1)}$$

where

$$\begin{aligned} \boldsymbol{E} &= [\, \mathbf{e}_{c1} \;\; \mathbf{e}_{c2} \; \ldots \; \mathbf{e}_{cM} \,] , \\ \boldsymbol{P} &= \mathrm{diag}(A_1^2, A_2^2, \ldots , A_M^2) , \\ \mathbf{e}_{ci} &= [\, 1 \;\; z_i \;\; z_i^2 \; \ldots \; z_i^{(L-1)} \,]^T , \\ z_i &= \exp[2\pi(-\text{œ}_i + jf_i)] , \\ a_i &= A_i e^{j\phi_i} , \end{aligned}$$

and the signal vector is given by

$$s(n) = \sum_{k=1}^{M} a_k z_k^n , \qquad 0 \le n \le N-1 . \qquad \text{(A1.2)}$$

Proof :

The theoretical autocorrelation function (ACF) for the process given by Equation (A1.2) is

$$\begin{aligned} r_{ss}(n,m) &= E[\, s(n)\, s^*(m) \,] \\ &= E\left[\left(\sum_{k=1}^{M} a_k z_k^n \right) \left(\sum_{i=1}^{M} a_i z_i^m \right)^* \right] \\ &= E\left[\sum_{k=1}^{M} \sum_{i=1}^{M} A_k A_i e^{j(\phi_k - \phi_i)} e^{-2\pi(\text{œ}_k n + \text{œ}_i m)} e^{j2\pi(f_k n - f_i m)} \right] . \end{aligned} \qquad \text{(A1.3)}$$

Assuming that the phases are independent random variables uniformly distributed on $[0, 2\pi)$, that is, the probability density functions of the phases are

$$p(\phi_i) = \frac{1}{2\pi} ,$$

and hence,

$$\begin{aligned} E\left[e^{j(\phi_k - \phi_i)}\right] &= \int_0^{2\pi} \int_0^{2\pi} e^{j(\phi_k - \phi_i)}\, p(\phi_i)\, p(\phi_k)\, d\phi_i\, d\phi_k \\ &= \delta_{i,k} , \end{aligned}$$

where $\delta_{i,k}$ denotes the Kronecker delta function. This means that the ACF becomes

$$\begin{aligned} r_{ss}(n,m) &= E\left[\sum_{i=1}^{M} A_i^2\, e^{-2\pi\sigma_i(n+m)}\, e^{j2\pi f_i(n-m)}\right] \\ &= \sum_{i=1}^{M} A_i^2\, z_i^n\, z_i^{*\,m} . \end{aligned}$$

Therefore, the L x L autocorrelation matrix is given by

$$\boldsymbol{R}_{ss} = \sum_{i=1}^{M} A_i^2\, \mathbf{e}_{ci}\, \mathbf{e}_{ci}^H .$$

Notice that the process is NOT wide-sense stationary, instead it is nonstationary.

Appendix 2 Proof of generalized Pisarenko property for damped sinusoids

It is proven in this appendix that the eigenvector associated with the minimum eigenvalues of an L x L Hermitian autocorrelation matrix for M complex exponentially damped sinusoids in white Gaussian noise described by Equation (1.1.1) has its zeros at the positions $z_i = \exp[2\pi(-\text{œ}_i + jf_i)]$, for i = 1, 2, . . . , M, assuming all the A_i's are nonzero and the frequencies to be distinct.

The theoretical total autocorrelation matrix can be decomposed as a sum of signal autocorrelation and noise autocorrelation matrices as follows :

$$\boldsymbol{R}_{yy} = \boldsymbol{R}_{ss} + 2\sigma_w^2 \boldsymbol{I}\,, \qquad \text{(A2.1)}$$

$$= \sum_{i=1}^{M} (\lambda_i + 2\sigma_w^2)\mathbf{v}_i \mathbf{v}_i^H + 2\sigma_w^2 \sum_{i=M+1}^{L} \mathbf{v}_i \mathbf{v}_i^H\,, \qquad \text{(A2.2)}$$

where the $\mathbf{v}_i$'s are orthonormal. Consider an arbitrary vector which is a linear combination of the eigenvectors in noise subspace, that is,

$$\mathbf{g} = \sum_{i=M+1}^{L} \beta_i \mathbf{v}_i\,, \qquad \text{(A2.3)}$$

where $\mathbf{g} = [\, g_0 \; g_1 \; \ldots . \; g_{L-1} \,]^T$ and the β_i's are some arbitrary constants. Then, together with Equation (A1.1), we have

$$\boldsymbol{R}_{ss}\,\mathbf{g} = \boldsymbol{E}\boldsymbol{P}\boldsymbol{E}^H\,\mathbf{g} = \mathbf{0}\,.$$

Premultiplying by $\mathbf{g}^H$ yields

$$\mathbf{g}^H \boldsymbol{E}\boldsymbol{P}\boldsymbol{E}^H \mathbf{g} = (\boldsymbol{E}^H \mathbf{g})^H \boldsymbol{P}\, (\boldsymbol{E}^H \mathbf{g}) = 0\,.$$

Since $\boldsymbol{P}$ is positive definite, it follows that $\boldsymbol{E}^H \mathbf{g} = \mathbf{0}$,

or $\mathbf{e}_{ci}^H \mathbf{g} = 0,$ for i = 1, 2, . . . , M,

or $\sum_{n=0}^{L-1} g_n \exp[2\pi(-\text{œ}_i - jf_i)n] = 0,$ for i = 1, 2, . . . , M, (A2.4)

which is the desired result.

Appendix 3 Definition of NMR spectroscopy frequency in units of "ppm"

$$\text{frequency [ppm]} = \frac{\text{frequency [Hz]}}{(\text{spectrometer frequency} + \text{offset frequency} * 10^{-6})\ [\text{MHz}]}$$

$$= \frac{\text{normalized frequency} * \text{spectral width [Hz]}}{(\text{spectrometer frequency} + \text{offset frequency} * 10^{-6})\ [\text{MHz}]}$$

$$= \frac{\text{normalized frequency} * \text{SW}}{\text{SF} + (\text{O1} * 10^{-6})} ,$$

where

SW = spectral width [Hz] = sampling frequency [Hz] ,

SF = spectrometer frequency [MHz] ,

O1 = offset frequency [Hz] .

Appendix 4 First order approximation of the discrete Fourier transform of complex damped sinusoid

Let $s(n\Delta t)$ be the complex exponentially damped sinusoid given as

$$s(n\Delta t) = A\, e^{j\phi}\, e^{2\pi(-\alpha+jf)n\Delta t} \,. \tag{A4.1}$$

Define the discrete Fourier transform pair as

$$S(mF) = \Delta t \sum_{n=0}^{N-1} s(n\Delta t)\, e^{-j2\pi nm/N}, \text{ where } F\Delta t = \frac{1}{N}; \tag{A4.2}$$

$$s(n\Delta t) = \frac{1}{N\Delta t} \sum_{n=0}^{N-1} S(mF)\, e^{j2\pi nm/N} \,. \tag{A4.3}$$

Substituting Equation (A4.1) in Equation (A4.2), we have

$$\begin{aligned}
S(mF) &= \Delta t \sum_{n=0}^{N-1} A\, e^{j\phi}\, e^{2\pi[-\alpha n\Delta t + j(f\Delta t - \frac{m}{N})n]} \\
&= \Delta t\, A\, e^{j\phi} \sum_{n=0}^{N-1} e^{2\pi[-\alpha n\Delta t + j(f\Delta t - \frac{m}{N})n]} \\
&= \Delta t\, A\, e^{j\phi}\, \frac{1 - e^{2\pi[-\alpha N\Delta t + j(f\Delta t - \frac{m}{N})N]}}{1 - e^{[-2\pi\alpha\Delta t + j2\pi(f\Delta t - \frac{m}{N})]}} \qquad \text{if } \alpha \neq 0 \text{ and } f \neq 0 \\
&\approx \Delta t\, A\, e^{j\phi}\, \frac{1}{1 - \{1 - 2\pi\alpha\Delta t + j2\pi(f\Delta t - \frac{m}{N})\}} \qquad \text{if } \alpha N\Delta t << 0 \\
&= \frac{A e^{j\phi}}{2\pi\alpha - j2\pi(f - \frac{m}{N\Delta t})}
\end{aligned}$$

The second last approximation was obtained from the following approximation :

$$e^{x} = 1 + x + \frac{x^2}{2} + \ldots\ldots$$

To show Equation (A4.3) from Equation (A4.2), substitute Equation (A4.2) into Equation (A4.3), i.e.,

$$s(n\Delta t) = \frac{1}{N\Delta t} \sum_{m=0}^{N-1} \Delta t \sum_{k=0}^{N-1} s(k\Delta t)\, e^{-j2\pi(k-n)m/N}$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} s(k\Delta t) \sum_{m=0}^{N-1} e^{-j2\pi(k-n)m/N}$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} s(k\Delta t) * \delta_{k,n}$$

$$= \frac{1}{N}\, s(n\Delta t)\; N$$

$$= s(n\Delta t) \;,$$

where $\delta_{k,n}$ is the Kronecker delta function.

Appendix 5 Amplitude response of window function in decibel scale

In studying the amplitude frequency response of window functions, it is desirable to employ a decibel scale. In each case, the decibel response is normalized with respect to the Fourier transform of the time domain window function evaluated at frequency zero.

Let W(f) represent the Fourier transform of a time domain window function w(t), and let W(0) represent the dc value of the transform. The amplitude response for the window function in decibel scale is defined by

$$W_{dB}(f) = 20 \log_{10} \frac{|W(f)|}{W(0)} .$$

Appendix 6 Proof of Theorem 3.2.1

The pseudo-inverse $\boldsymbol{A}^{+}$ of a matrix $\boldsymbol{A} \in \mathrm{C}^{n \times m}$ is unique.

Proof:

Suppose $\boldsymbol{B} \in \mathrm{C}^{n \times m}$, $\boldsymbol{C} \in \mathrm{C}^{n \times m}$, and $\boldsymbol{B} \neq \boldsymbol{C}$, but both are pseudo-inverses of $\boldsymbol{A}$, then

$$\begin{aligned} B \quad &= B\underline{A}\underline{B} = BB^{\mathrm{H}}\underline{A}^{\mathrm{H}} = BB^{\mathrm{H}}(ACA)^{\mathrm{H}} = BB^{\mathrm{H}}A^{\mathrm{H}}\,\underline{C}^{\mathrm{H}}\underline{A}^{\mathrm{H}} \\ &= B\underline{B}^{\mathrm{H}}\underline{A}^{\mathrm{H}}\,AC = \underline{BAB}AC = B\underline{A}C = BA\underline{CA}C \\ &= \underline{BA}A^{\mathrm{H}}C^{\mathrm{H}}\,C = \underline{A}^{\mathrm{H}}\underline{B}^{\mathrm{H}}\,\underline{A}^{\mathrm{H}}C^{\mathrm{H}}\,C \\ &= \underline{A}^{\mathrm{H}}\underline{C}^{\mathrm{H}}\,C = CAC = C. \end{aligned}$$

An alternative proof of the uniqueness of $\boldsymbol{A}^{+}$ can be found in Boullion and Odell, 1971, p.3.

Appendix 7 Proof of Theorem 3.2.2

The matrix $\boldsymbol{G} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{+}$ is a pseudo-inverse of $\boldsymbol{A}$.

Proof: First consider $\boldsymbol{G} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}}$. We have

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{G}\boldsymbol{A} &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} \\
&= \boldsymbol{A}\boldsymbol{A}^{+}(\boldsymbol{A}^{\mathrm{H}})^{+}(\boldsymbol{A}^{\mathrm{H}})\boldsymbol{A} \\
&= (\boldsymbol{A}\boldsymbol{A}^{+})(\boldsymbol{A}\boldsymbol{A}^{+})^{\mathrm{H}}\boldsymbol{A} \\
&= (\boldsymbol{A}\boldsymbol{A}^{+})(\boldsymbol{A}\boldsymbol{A}^{+})\boldsymbol{A} \\
&= \boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{A} \\
&= \boldsymbol{A}\,.
\end{aligned}$$

$$\boldsymbol{G}\boldsymbol{A}\boldsymbol{G} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{G}\,.$$

$$(\boldsymbol{G}\boldsymbol{A})^{\mathrm{H}} = [(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})]^{\mathrm{H}} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{G}\boldsymbol{A}\,.$$

$$\begin{aligned}
(\boldsymbol{A}\boldsymbol{G})^{\mathrm{H}} &= [\boldsymbol{A}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}}]^{\mathrm{H}} \\
&= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}\boldsymbol{G}\,.
\end{aligned}$$

The identity $\boldsymbol{G} = \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{+}$ can be proved similarly.

Appendix 8 Proof of Theorem 3.2.4

Let $\boldsymbol{A} \in \mathrm{C}^{m \times n}$, $\mathbf{b} \in \mathrm{C}^m$ and $\mathbf{x}_{\mathrm{LS}} = \boldsymbol{A}^+\mathbf{b}$. Then each $\mathbf{x} \in \mathrm{C}^n$, $\mathbf{x} \neq \mathbf{x}_{\mathrm{LS}}$, satisfies one of the two conditions :

(a) $\| \boldsymbol{A}\,\mathbf{x}_{\mathrm{LS}} - \mathbf{b} \|_2 < \| \boldsymbol{A}\,\mathbf{x} - \mathbf{b} \|_2$ or

(b) $\| \boldsymbol{A}\,\mathbf{x}_{\mathrm{LS}} - \mathbf{b} \|_2 = \| \boldsymbol{A}\,\mathbf{x} - \mathbf{b} \|_2$ and $\| \mathbf{x}_{\mathrm{LS}} \|_2 < \| \mathbf{x} \|_2$.

Proof :

(a) For any $\mathbf{x} \in \mathrm{C}^n$, write

$$\begin{aligned} \boldsymbol{A}\,\mathbf{x} - \mathbf{b} &= \boldsymbol{A}\,\mathbf{x} - \boldsymbol{A}\,\boldsymbol{A}^+\mathbf{b} + \boldsymbol{A}\,\boldsymbol{A}^+\mathbf{b} - \mathbf{b} \\ &= \boldsymbol{A}\,(\mathbf{x} - \boldsymbol{A}^+\mathbf{b}) - (\boldsymbol{I} - \boldsymbol{A}\,\boldsymbol{A}^+)\,\mathbf{b} \\ &= \boldsymbol{A}\,(\mathbf{x} - \mathbf{x}_{\mathrm{LS}}) - (\boldsymbol{I} - \boldsymbol{A}\,\boldsymbol{A}^+)\,\mathbf{b} \ . \end{aligned}$$

Therefore,

$$\begin{aligned} \| \boldsymbol{A}\,\mathbf{x} - \mathbf{b} \|_2 &= \| \boldsymbol{A}\,(\mathbf{x} - \mathbf{x}_{\mathrm{LS}}) \|_2 + \| (\boldsymbol{I} - \boldsymbol{A}\,\boldsymbol{A}^+)\,\mathbf{b} \|_2 \\ &= \| \boldsymbol{A}\,(\mathbf{x} - \mathbf{x}_{\mathrm{LS}}) \|_2 + \| \boldsymbol{A}\,\mathbf{x}_{\mathrm{LS}} - \mathbf{b} \|_2 \\ &\geq \| \boldsymbol{A}\,\mathbf{x}_{\mathrm{LS}} - \mathbf{b} \|_2 , \end{aligned}$$

with equality only if $\boldsymbol{A}\,\mathbf{x} = \boldsymbol{A}\,\mathbf{x}_{\mathrm{LS}}$.

(b) In the case of $\boldsymbol{A}\,\mathbf{x} = \boldsymbol{A}\,\mathbf{x}_{\mathrm{LS}}$, we have, since $\mathbf{x}_{\mathrm{LS}} = \boldsymbol{A}^+\mathbf{b}$,

$$\boldsymbol{A}^+\boldsymbol{A}\,\mathbf{x} = \boldsymbol{A}^+\boldsymbol{A}\,\mathbf{x}_{\mathrm{LS}} = \boldsymbol{A}^+\boldsymbol{A}\,\boldsymbol{A}^+\mathbf{b} = \boldsymbol{A}^+\mathbf{b} = \mathbf{x}_{\mathrm{LS}} \ .$$

Thus, we may write

$$\mathbf{x} = \mathbf{x}_{LS} + (\mathbf{x} - \mathbf{x}_{LS}) = A^+\mathbf{b} + (I - A^+A)\mathbf{x} .$$

Therefore,

$$\begin{aligned} \| \mathbf{x} \|_2 &= \| A^+\mathbf{b} \|_2 + \| (I - A^+A)\mathbf{x} \|_2 \\ &= \| \mathbf{x}_{LS} \|_2 + \| \mathbf{x} - \mathbf{x}_{LS} \|_2 \\ &\geq \| \mathbf{x}_{LS} \|_2 , \end{aligned}$$

with equality if and only if $\mathbf{x} = \mathbf{x}_{LS}$.

Theorem 3.2.3 can be restated as follows :

Let $A \in \mathbb{C}^{m \times n}$, $\mathbf{b} \in \mathbb{C}^m$. Then, amongst the least-squares solutions of $A\mathbf{x} = \mathbf{b}$, $A^+\mathbf{b}$ is the one of minimum 2-norm. Conversely, if $G \in \mathbb{C}^{n \times m}$ has the property that, for all $\mathbf{b}$, $G\mathbf{b}$ is the minimum 2-norm least-squares solution of $A\mathbf{x} = \mathbf{b}$, then $G = A^+$.

Appendix 9 Proof of Theorem 4.2.1.1

Let $s(n) = \sum_{k=1}^{M} a_k e^{s_k n}$, $n = 0, 1, \dots, N-1$, where a_k and s_k are unknown complex numbers, $s_k \neq s_i$ for $k \neq i$, and

$$\boldsymbol{A}'_F = \begin{bmatrix} s(L) & s(L-1) & \dots & s(0) \\ s(L+1) & s(L) & \dots & s(1) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ s(N-1) & s(N-2) & \dots & s(N-L-1) \end{bmatrix} = [\, \boldsymbol{h}_F, \boldsymbol{A}_F \,] \ .$$

If the coefficient vector $\mathbf{g}'_F = [\, g_{F,0}, g_{F,1}, g_{F,2}, \dots, g_{F,L}]^T$ satisfies the homogeneous equation $\boldsymbol{A}'_F \mathbf{g}'_F = \mathbf{0}$, and if L satisfies the inequality $M \leq L \leq N-M$, then $G_F(z) = \sum_{k=0}^{L} g_{F,k}\, z^{-k}$ has M of its L zeros at e^{s_k}, for $k = 1, 2, \dots, M$.

Proof:

If s(n) is given as described in the theorem, then one can show that $\boldsymbol{A}'_F$ can be rewritten as

$$\boldsymbol{A}'_F = \boldsymbol{Z}_{FL}\, \boldsymbol{Z}'_{FR} \ ,$$

where

$$\boldsymbol{Z}_{FL} = \begin{bmatrix} \Psi(1,L) & \Psi(2,L) & \dots & \Psi(M,L) \\ \Psi(1,L+1) & \Psi(2,L+1) & \dots & \Psi(M,L+1) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \Psi(1,N-1) & \Psi(2,N-1) & \dots & \Psi(M,N-1) \end{bmatrix} , \qquad (N-L) \times M$$

$$\Psi(k,h) = |a_k|\, e^{s_k h + j\phi_k} ,$$

$$Z'_{FR} = \begin{bmatrix} 1 & e^{-s_1} & e^{-2s_1} & \dots & e^{-Ls_1} \\ 1 & e^{-s_2} & e^{-2s_2} & \dots & e^{-Ls_2} \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ 1 & e^{-s_M} & e^{-2s_M} & \dots & e^{-Ls_M} \end{bmatrix} . \qquad M \times (L+1)$$

We know that $A_F\,\mathbf{g} = \mathbf{h}_F$ has a nontrivial solution if and only if $\mathrm{rank}(A_F) = \mathrm{rank}(A_F, \mathbf{h}_F)$. Suppose $s_i \neq s_j$ for $i \neq j$. By Sylvester's law (Marcus and Minc, 1964, p.28),

$$\mathrm{rank}(Z_{FL}) + \mathrm{rank}(Z'_{FR}) - M \le \mathrm{rank}(Z_{FL}Z'_{FR}) \le \min \{ \mathrm{rank}(Z_{FL}), \mathrm{rank}(Z'_{FR}) \}$$

$$\therefore (M + M - M) \le \mathrm{rank}(A'_F) \le \min \{ M, M \} \qquad \text{if } (N - L) \ge M \text{ and } M \le (L + 1).$$

$$\therefore \mathrm{rank}(A'_F) = M \qquad \text{if } M - 1 \le L \le N - M .$$

Similarly,

$$\mathrm{rank}(A_F) = M \qquad \text{if } M \le L \le N - M .$$

Therefore,

$$\mathrm{rank}(A_F) = \mathrm{rank}(A'_F) = M \qquad \text{if } M \le L \le N - M . \qquad (A9.1)$$

Equation (A9.1) means that the row vectors of A'_F are linear combinations of the M row vectors of Z'_{FR} .

Suppose $\mathrm{rank}(A'_F) = M$. Let $\mathbf{f}_F(k)^H$ be the k-th row vector of Z'_{FR} , i.e.,

$$\mathbf{f}_F(k)^H = (1, e^{-s_k}, e^{-2s_k}, \dots, e^{-Ls_k}) , \qquad k = 1, 2, \dots, M.$$

Then, the i-th row of $A'_F\,\mathbf{g}'_F$ is

$$\sum_{k=1}^{M} a_{ik}\, \mathbf{f}_F(k)^H\, \mathbf{g}'_F = 0$$

$$\Rightarrow \quad \sum_{k=1}^{M} |a_k|\, e^{s_k (L+i) + j\phi_k}\, f_F(k)^H g'_F = 0$$

$$\Rightarrow \quad |a_1|\, e^{s_1 (L+i) + j\phi_1} \,(g_{F,0} + g_{F,1} e^{-s_1} + g_{F,2} e^{-2s_1} + \ldots + g_{F,L} e^{-Ls_1})$$

$$+ \; |a_2|\, e^{s_2 (L+i) + j\phi_2} \,(g_{F,0} + g_{F,1} e^{-s_2} + g_{F,2} e^{-2s_2} + \ldots + g_{F,L} e^{-Ls_2})$$

$$\vdots$$

$$+ \; |a_M|\, e^{s_M (L+i) + j\phi_M} \,(g_{F,0} + g_{F,1} e^{-s_M} + g_{F,2} e^{-2s_M} + \ldots + g_{F,L} e^{-Ls_M}) = 0 \,.$$

In other words, the polynomial

$$G_F(z) = \sum_{k=0}^{L} g_{F,k}\, z^{-k}$$

has M of its zeros at e^{s_k} for $k = 1, 2, \ldots, M$.

Appendix 10 Proof of Theorem 4.2.1.2

Let $s(n) = \sum_{k=1}^{M} a_k e^{s_k n}$, n = 0, 1, ... , N-1, where a_k is the complex amplitude with unknown magnitude A_k and phase ϕ_k; $s_k/2\pi$ is the complex frequency with unknown normalized damping rate α_k and normalized frequency f_k. $s_i \neq s_k$ for $i \neq k$. M is the number of complex sinusoids. If a coefficient vector $\mathbf{g}'_B = [\, g_{B,0}, g_{B,1}, g_{B,2}, \ldots, g_{B,L}]^T$ satisfies the homogeneous equation $\boldsymbol{A}'_B\, \mathbf{g}'_B = \mathbf{0}$, and if L satisfies the inequality $M \leq L \leq N\text{-}M$, the polynomial $G_B(z) = \sum_{k=0}^{L} g_{B,k}\, z^{-k}$ has zeros at $e^{-s_k^*}$, for k = 1, 2, ..., M, where

$$\boldsymbol{A}'_B = \begin{bmatrix} s^*(0) & s^*(1) & \ldots & s^*(L) \\ s^*(1) & s^*(2) & \ldots & s^*(L+1) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ s^*(N\text{-}L\text{-}1) & s^*(N\text{-}L) & \ldots & s^*(N\text{-}1) \end{bmatrix} = [\mathbf{h}_B, \boldsymbol{A}_B]\,. \qquad (4.2.1.5)$$

Proof: The proof is the same as that of Theorem 4.2.1.1.

Rewrite $\boldsymbol{A}'_B$ as follows

$$\boldsymbol{A}'_B = \boldsymbol{Z}_{BL}\, \boldsymbol{Z}'_{BR}\ ,$$

with

$$\boldsymbol{Z}_{BL} = \begin{bmatrix} \Psi(1,0) & \Psi(2,0) & \ldots & \Psi(M,0) \\ \Psi(1,1) & \Psi(2,1) & \ldots & \Psi(M,1) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \Psi(1,N\text{-}L\text{-}1) & \Psi(2,N\text{-}L\text{-}1) & \ldots & \Psi(M,N\text{-}L\text{-}1) \end{bmatrix}, \quad (N\text{-}L) \times M\ ,$$

$$\Psi(k,h) = |a_k|\, e^{h s_k^* - j \phi_k},$$

$$\boldsymbol{Z}'_{BR} = \begin{bmatrix} 1 & e^{s_1^*} & e^{2s_1^*} & \dots & e^{Ls_1^*} \\ 1 & e^{s_2^*} & e^{2s_2^*} & \dots & e^{Ls_2^*} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & e^{s_M^*} & e^{2s_M^*} & \dots & e^{Ls_M^*} \end{bmatrix}, \qquad M \times (L+1)$$

and

$$\boldsymbol{A}_B = \boldsymbol{Z}_{BL}\, \boldsymbol{Z}_{BR} \,,$$

where

$$\boldsymbol{Z}_{BR} = \begin{bmatrix} e^{s_1^*} & e^{2s_1^*} & \dots & e^{Ls_1^*} \\ e^{s_2^*} & e^{2s_2^*} & \dots & e^{Ls_2^*} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ e^{s_M^*} & e^{2s_M^*} & \dots & e^{Ls_M^*} \end{bmatrix}. \qquad M \times L$$

Again by Sylvester's law (Marcus and Minc, 1964, pp. 28), we have

$$\text{rank}\,(\boldsymbol{A}'_B) = M \qquad \text{if } M - 1 \le L \le N - M\,,$$

$$\text{rank}(\boldsymbol{A}_B) = M \qquad \text{if } M \le L \le N - M\,.$$

In order for $\boldsymbol{A}_B\, \mathbf{g}_B = \mathbf{h}_B$ to have a nontrivial solution, we require

$$\text{rank}(\boldsymbol{A}'_B) = \text{rank}(\boldsymbol{A}_B) = M\,,$$

and this can happen if $M \le L \le N - M\,.$

This means that each row vector of $\boldsymbol{A}'_B$ is a linear combination of the M row vectors of $\boldsymbol{Z}'_{BR}$.

If rank($\boldsymbol{A}'_B$) = M , let

$$f_k^H = (1, e^{s_k^*}, e^{2s_k^*}, \ldots, e^{Ls_k^*}), \qquad k = 1, 2, \ldots, M.$$

Then the i-th row vector of $\boldsymbol{A}'_B\, \mathbf{g}'_B$ can be written as,

$$\sum_{k=1}^{M} a_{ik}\, f_k^H\, g'_B = 0$$

$$\Rightarrow \sum_{k=1}^{M} |a_k|\, e^{is_k^* - j\phi_k}\, f_k^H\, g'_B = 0$$

$$\Rightarrow |a_1|\, e^{is_1^* - j\phi_1} \left(g_{B,0} + g_{B,1}e^{s_1^*} + g_{B,2}e^{2s_1^*} + \ldots + g_{B,L}e^{Ls_1^*} \right)$$

$$+ |a_2|\, e^{is_2^* - j\phi_2} \left(g_{B,0} + g_{B,1}e^{s_2^*} + g_{B,2}e^{2s_2^*} + \ldots + g_{B,L}e^{Ls_2^*} \right)$$

$$\vdots$$

$$+ |a_M|\, e^{is_M^* - j\phi_M} \left(g_{B,0} + g_{B,1}e^{s_M^*} + g_{B,2}e^{2s_M^*} + \ldots + g_{B,L}e^{Ls_M^*} \right) = 0 .$$

And hence the polynomial

$$G_B(z) = \sum_{k=0}^{L} g_{B,k}\, z^{-k}$$

has M of its zeros at $e^{-s_k^*}$, for k = 1, 2, ... , M.

Appendix 11 Proof of Theorem 4.2.1.3

Let $G(z) = 1 + \sum_{k=1}^{L} g_k z^{-k}$, $B(z) = \sum_{k=0}^{M} b_k z^{-k}$, $C(z) = \sum_{k=0}^{L-M} c_k z^{-k}$, and $G(z) = B(z)C(z)$. If $Q = 1 + |g_1|^2 + |g_2|^2 + \ldots + |g_L|^2$ is minimum, and $B(z)$ is given, then the roots of $C(z)$ must be inside the unit circle (see e.g., Lang and McClellan, 1979; Pakula and Kay, 1983; Stoica and Nehorai, 1987).

Proof:

Before we prove the theorem, we will try to find out the coefficients of C(z) if B(z) is given and Q is minimized. To do this, from the definition of G(z), we have

$$G(z) = B(z)\,C(z)\,. \tag{A11.1}$$

$$\begin{aligned}
\therefore\; 1 + \sum_{k=1}^{L} g_k z^{-k} &= \left(\sum_{k=0}^{M} b_k z^{-k}\right)\left(\sum_{j=0}^{L-M} c_j z^{-j}\right) \\
&= \left(b_0 + \sum_{k=1}^{M} b_k z^{-k}\right)\left(c_0 + \sum_{j=1}^{L-M} c_j z^{-j}\right) \\
&= b_0 c_0 + c_0 \sum_{k=1}^{M} b_k z^{-k} + b_0 \sum_{j=1}^{L-M} c_j z^{-j} + \sum_{k=1}^{M} \sum_{m=k+1}^{L-M+k} b_k c_{m-k} z^{-m} \\
&= b_0 c_0 + c_0 \sum_{k=1}^{M} b_k z^{-k} + \sum_{k=0}^{0} \sum_{m=k+1}^{L-M+k} b_k c_{m-k} z^{-m} \\
&\quad + \sum_{k=1}^{M} \sum_{m=k+1}^{L-M+k} b_k c_{m-k} z^{-m} \\
&= b_0 c_0 + c_0 \sum_{k=1}^{M} b_k z^{-k} + \sum_{k=0}^{M} \sum_{m=k+1}^{L-M+k} b_k c_{m-k} z^{-m}\,.
\end{aligned}$$

Equating the coefficients for different powers, we have

$$1 = b_0 c_0 \quad \Rightarrow \quad c_0 = \frac{1}{b_0}, \tag{A11.2}$$

and

$$\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ g_L \end{bmatrix} = c_0 \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \cdot \\ b_M \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} b_0 & 0 & \dots & 0 \\ \cdot & b_0 & \cdot & \cdot \\ \cdot & & \cdot & 0 \\ b_M & & & b_0 \\ 0 & b_M & & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & b_M \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \cdot \\ \cdot \\ c_{L\text{-}M} \end{bmatrix} .$$

$$\therefore \quad \mathbf{g} = c_0\mathbf{b} + \boldsymbol{B}_{L\text{-}M}\mathbf{c}$$

$$\therefore \quad \mathbf{g}^H\mathbf{g} = \left(c_0\mathbf{b} + \boldsymbol{B}_{L\text{-}M}\mathbf{c}\right)^H \left(c_0\mathbf{b} + \boldsymbol{B}_{L\text{-}M}\mathbf{c}\right)$$

$$= \left(c_0{}^*\mathbf{b}^H + \mathbf{c}^H\boldsymbol{B}_{L\text{-}M}^H\right)\left(c_0\mathbf{b} + \boldsymbol{B}_{L\text{-}M}\mathbf{c}\right)$$

$$= |c_0|^2\,\mathbf{b}^H\mathbf{b} + c_0{}^*\mathbf{b}^H\boldsymbol{B}_{L\text{-}M}\mathbf{c} + c_0\mathbf{c}^H\boldsymbol{B}_{L\text{-}M}^H\mathbf{b} + \mathbf{c}^H\boldsymbol{B}_{L\text{-}M}^H\boldsymbol{B}_{L\text{-}M}\mathbf{c} .$$

$$\frac{\partial}{\partial \mathbf{c}}(\mathbf{g}^H\mathbf{g}) = 2c_0\boldsymbol{B}_{L\text{-}M}^H\mathbf{b} + 2\boldsymbol{B}_{L\text{-}M}^H\boldsymbol{B}_{L\text{-}M}\mathbf{c} = 0 \text{ for } \mathbf{g}^H\mathbf{g} \text{ to be minimized.}$$

$$\therefore \quad \mathbf{g}^H\mathbf{g} \text{ is minimum if } \quad \boldsymbol{B}_{L\text{-}M}^H\boldsymbol{B}_{L\text{-}M}\mathbf{c} = -c_0\boldsymbol{B}_{L\text{-}M}^H\mathbf{b}. \tag{A11.3}$$

Combining Equations (A11.2) and (A11.3), we have

$$\boldsymbol{R}_{L\text{-}M+1}\mathbf{c}' = \boldsymbol{B}_{L\text{-}M+1}^H\boldsymbol{B}_{L\text{-}M+1}\mathbf{c}' = c_0\begin{bmatrix} |b_0|^2 \\ \mathbf{0} \end{bmatrix}, \tag{A11.4}$$

where

$$\boldsymbol{B}_{L\text{-}M+1} = \begin{bmatrix} b_0 & \mathbf{0}^H \\ \mathbf{b} & \boldsymbol{B}_{L\text{-}M} \end{bmatrix}, \tag{A11.5}$$

$$\mathbf{c}' = [\ c_0,\ \mathbf{c}\ ]^T , \tag{A11.6}$$

$$\mathbf{b} = [\ b_1, b_2, \dots, b_M, 0, \dots\dots, 0\ ]^T . \tag{A11.7}$$

Equation (A11.4) can be shown as follows :

$$\boldsymbol{B}^{H}_{L\text{-}M+1}\,\boldsymbol{B}_{L\text{-}M+1}\,\mathbf{c}' = \begin{bmatrix} b_0^* & \mathbf{b}^H \\ 0 & \boldsymbol{B}^{H}_{L\text{-}M} \end{bmatrix} \begin{bmatrix} b_0 & \mathbf{0}^H \\ \mathbf{b} & \boldsymbol{B}_{L\text{-}M} \end{bmatrix} \begin{bmatrix} c_0 \\ \mathbf{c} \end{bmatrix}$$

$$= \begin{bmatrix} c_0 b_0^* b_0 + c_0\mathbf{b}^H\mathbf{b} + \mathbf{b}^H\boldsymbol{B}_{L\text{-}M}\mathbf{c} \\ c_0\boldsymbol{B}^{H}_{L\text{-}M}\mathbf{b} + \boldsymbol{B}^{H}_{L\text{-}M}\boldsymbol{B}_{L\text{-}M}\mathbf{c} \end{bmatrix}. \qquad \text{(A11.8)}$$

From Equation (A11.3) we have

$$\boldsymbol{B}^{H}_{L\text{-}M}\boldsymbol{B}_{L\text{-}M}\mathbf{c} = -c_0\boldsymbol{B}^{H}_{L\text{-}M}\mathbf{b}$$

$$\therefore \quad [\boldsymbol{B}^{H}_{L\text{-}M}]^{+}\boldsymbol{B}^{H}_{L\text{-}M}\boldsymbol{B}_{L\text{-}M}\mathbf{c} = -c_0[\boldsymbol{B}^{H}_{L\text{-}M}]^{+}\boldsymbol{B}^{H}_{L\text{-}M}\mathbf{b}$$

$$\therefore \quad \boldsymbol{B}_{L\text{-}M}\mathbf{c} = -c_0\mathbf{b}\,. \qquad \text{(A11.9)}$$

Equation (A11.9) was obtained by using the full rank of $\boldsymbol{B}^{H}_{L\text{-}M}$. Substituting Equation (A11.9) into Equation (A11.8), we can prove Equation (A11.4).

Now we are ready to prove Theorem 4.2.1.3. Let us introduce the following companion matrix :

$$\boldsymbol{C} = \begin{bmatrix} -c_1/c_0 & 1 & 0 & \dots & 0 \\ \cdot & 0 & \cdot\,\cdot & & \\ \cdot & & & & \cdot \\ -c_{L\text{-}M+1}/c_0 & 0 & & & 1 \\ -c_{L\text{-}M}/c_0 & 0 & \dots & \dots & 0 \end{bmatrix} = \begin{bmatrix} -\dfrac{\mathbf{c}}{c_0} & \begin{matrix} \boldsymbol{I} \\ \mathbf{0}^H \end{matrix} \end{bmatrix}. \qquad \text{(A11.10)}$$

The roots of $C(z) = \sum_{k=0}^{L\text{-}M} c_k z^{-k}$ are equal to the eigenvalues of $\boldsymbol{C}$ (see e.g., Usmani, 1987, pp. 138). Let λ denote an arbitrary eigenvalue of $\boldsymbol{C}$ and $\mathbf{u} = [1, u_2, \dots, u_{L\text{-}M}]^T \neq \mathbf{0}$ the associated eigenvector

$$\boldsymbol{C}\,\mathbf{u} = \lambda\,\mathbf{u}\,,$$

or equivalently,

$$\begin{bmatrix} \mathbf{u} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{\mathbf{c}}{c_0} \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ \mathbf{u} \end{bmatrix}. \qquad \text{(A11.11)}$$

It follows from Equations (A11.3), (A11.4), (A11.5), (A11.7) and the Toeplitz structure of $\boldsymbol{R}_{L-M+1}$ that

$$\begin{aligned}
\mathbf{u}^{H} \boldsymbol{R}_{L-M} \mathbf{u} &= [\, \mathbf{u}^{H} \ 0 \,] \boldsymbol{R}_{L-M} \begin{bmatrix} \mathbf{u} \\ 0 \end{bmatrix} \\
&= \left\{ \begin{bmatrix} 1 & \frac{\mathbf{c}^{H}}{c_0^{*}} \end{bmatrix} + \lambda^{*} [\, 0 \ \mathbf{u}^{H} \,] \right\} \boldsymbol{R}_{L-M+1} \left\{ \begin{bmatrix} 1 \\ \frac{\mathbf{c}}{c_0} \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ \mathbf{u} \end{bmatrix} \right\} \\
&= \left\{ \begin{bmatrix} 1 & \frac{\mathbf{c}^{H}}{c_0^{*}} \end{bmatrix} + \lambda^{*} [\, 0 \ \mathbf{u}^{H} \,] \right\} \left\{ \begin{bmatrix} |b_0|^2 \\ 0 \end{bmatrix} + \lambda \boldsymbol{R}_{L-M+1} \begin{bmatrix} 0 \\ \mathbf{u} \end{bmatrix} \right\} \\
&= |b_0|^2 + \lambda \begin{bmatrix} 1 & \frac{\mathbf{c}^{H}}{c_0^{*}} \end{bmatrix} \boldsymbol{R}_{L-M+1} \begin{bmatrix} 0 \\ \mathbf{u} \end{bmatrix} + \\
&\qquad |\lambda|^2 [\, 0 \ \mathbf{u}^{H} \,] \boldsymbol{R}_{L-M+1} \begin{bmatrix} 0 \\ \mathbf{u} \end{bmatrix} \\
&= |b_0|^2 + \lambda \begin{bmatrix} 1 & \frac{\mathbf{c}^{H}}{c_0^{*}} \end{bmatrix} \begin{bmatrix} |b_0|^2 & b^{H} \boldsymbol{B}_{L-M} \\ \boldsymbol{B}_{L-M}^{H} b & \boldsymbol{R}_{L-M} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{u} \end{bmatrix} + \\
&\qquad |\lambda|^2 [\, 0 \ \mathbf{u}^{H} \,] \begin{bmatrix} |b_0|^2 & b^{H} \boldsymbol{B}_{L-M} \\ \boldsymbol{B}_{L-M}^{H} b & \boldsymbol{R}_{L-M} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{u} \end{bmatrix} \\
&= |b_0|^2 + \lambda \left[b^{H} \boldsymbol{B}_{L-M} \mathbf{u} + \frac{\mathbf{c}^{H}}{c_0^{*}} \boldsymbol{R}_{L-M} \mathbf{u} \right] + |\lambda|^2 \mathbf{u}^{H} \boldsymbol{R}_{L-M} \mathbf{u} \\
&= |b_0|^2 + \lambda \left[b^{H} \boldsymbol{B}_{L-M} \mathbf{u} - \frac{c_0^{*}}{c_0^{*}} b^{H} \boldsymbol{B}_{L-M} \mathbf{u} \right] + |\lambda|^2 \mathbf{u}^{H} \boldsymbol{R}_{L-M} \mathbf{u} \\
&= |b_0|^2 + |\lambda|^2 \mathbf{u}^{H} \boldsymbol{R}_{L-M} \mathbf{u} \, .
\end{aligned}$$

$$\therefore \qquad |\lambda|^2 = 1 - \frac{|b_0|^2}{\mathbf{u}^{H} \boldsymbol{R}_{L-M} \mathbf{u}} \, . \qquad \text{(A11.12)}$$

We know that $\mathbf{u}^{H} \boldsymbol{R}_{L-M} \mathbf{u} \geq 0$ because

$$\mathbf{u}^{H} \boldsymbol{R}_{L-M} \mathbf{u} = \mathbf{u}^{H} \boldsymbol{B}_{L-M}^{H} \boldsymbol{B}_{L-M} \mathbf{u} = \| \boldsymbol{B}_{L-M} \mathbf{u} \|_2^2 \geq 0.$$

In addition, $\| \boldsymbol{B}_{L\text{-}M}\mathbf{u} \|_2^2 = 0$, if and only if $\boldsymbol{B}_{L\text{-}M}\mathbf{u} = \mathbf{0}$. However, because of the Toeplitz structure of $\boldsymbol{B}_{L\text{-}M}$,

$$\boldsymbol{B}_{L\text{-}M}\mathbf{u} = \begin{bmatrix} b_0 & 0 & \dots & 0 \\ \cdot & b_0 & \cdot & \cdot \\ \cdot & & \cdot & 0 \\ b_M & & & b_0 \\ 0 & b_M & & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & b_M \end{bmatrix} \begin{bmatrix} 1 \\ u_2 \\ u_3 \\ \cdot \\ \cdot \\ u_{L\text{-}M} \end{bmatrix} = \mathbf{0}$$

if and only if all b_i's are zeros. Therefore, we conclude that $\mathbf{u}^H\boldsymbol{R}_{L\text{-}M}\mathbf{u} > 0$, and so from Equation (A11.12), $|\lambda|^2$ is always less than one, hence, the zeros of C(z) are always within the unit circle.

Appendix 12 Proof of Theorem 4.2.2.1

Proof :

From Wedin (1973) we have

$$\hat{A}^+ - A^+ = -\hat{A}^+ \Delta A \, A^+ + (\hat{A}^H \hat{A})^+ \Delta A^H P_A^{\perp} + R_{\hat{A}}^{\perp} \Delta A^H (AA^H)^+ , \tag{A12.1}$$

where $P_A^{\perp} = I - AA^+$ is the projector onto the orthogonal complement of the column space of A ; $R_{\hat{A}}^{\perp} = I - \hat{A}^+\hat{A}$ is the projector onto the orthogonal complement of the row space of $\hat{A}$.

Let $\hat{\mathbf{v}}_0$ be any vector belonging to the space spanned by $\{ \hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_M \}$ and $\mathbf{u}_0$ be any vector belonging to the column space of A, then

$$\begin{aligned} P_A^{\perp}\mathbf{u}_0 &= 0 , \\ R_{\hat{A}}^{\perp}\hat{\mathbf{v}}_0 &= 0 . \end{aligned}$$

Thus, from Equation (A12.1) we have

$$\hat{\mathbf{v}}_0^H (\hat{A}^+ - A^+) \mathbf{u}_0 = -\hat{\mathbf{v}}_0^H \hat{A}^+ \Delta A \, A^+ \mathbf{u}_0 . \tag{A12.2}$$

However, by the SVD of $\hat{A}$, we have

$$\hat{A}^+ = \sum_{i=1}^{\hat{M}} \frac{1}{\hat{\sigma}_i} \hat{\mathbf{v}}_i \hat{\mathbf{u}}_i^H ,$$

and so,

$$\hat{A}_M^+ = \sum_{i=1}^{M} \frac{1}{\hat{\sigma}_i} \hat{\mathbf{v}}_i \hat{\mathbf{u}}_i^H .$$

Therefore,

$$\hat{\mathbf{v}}_0^H \hat{A}^+ = \hat{\mathbf{v}}_0^H \hat{A}_M^+ , \tag{A12.3}$$

and so, from Equation (A12.2),

$$\begin{aligned} -\hat{\mathbf{v}}_0^H \hat{A}^+ \Delta A \, A^+ \mathbf{u}_0 &= \hat{\mathbf{v}}_0^H (\hat{A}^+ - A^+) \mathbf{u}_0 \\ &= \hat{\mathbf{v}}_0^H (\hat{A}_M^+ - A^+) \mathbf{u}_0 \\ &= \hat{\mathbf{v}}_0^H \Delta A_M^+ \mathbf{u}_0 . \end{aligned} \tag{A12.4}$$

Let $\Delta\boldsymbol{A}$ approach zero, so we have

$$\begin{aligned} \hat{\boldsymbol{A}}^{+} &\rightarrow \boldsymbol{A}^{+}, \\ \hat{\mathbf{v}}_0{}^{\mathrm{H}} &\rightarrow \mathbf{v}_0{}^{\mathrm{H}}. \end{aligned}$$

and hence Equation (A12.4) becomes

$$\mathbf{v}_0{}^{\mathrm{H}}\, \Delta\boldsymbol{A}^{+}_{\mathrm{M}}\, \mathbf{u}_0 \quad = \quad - \mathbf{v}_0{}^{\mathrm{H}}\, \boldsymbol{A}^{+}\, \Delta\boldsymbol{A}\, \boldsymbol{A}^{+}\, \mathbf{u}_0 \,.$$

Appendix 13 Proof of Equation (4.2.2.26)

Proof :

First, we derive the first order perturbation of Equation (4.2.5) :

$$1+\sum_{k=1}^{L}\hat{g}_k\,\hat{z}^{-k} \;=\; 0 \qquad (4.2.5)$$

$$\therefore \quad 1+\sum_{k=1}^{L}(g_k+\Delta g_k)(z+\Delta z)^{-k} \;=\; 0$$

$$\therefore \quad 1+\sum_{k=1}^{L}(g_k+\Delta g_k)(z^{-k}+k\,z^{-k-1}\Delta z) \;\approx\; 0$$

$$\therefore \quad [1+\sum_{k=1}^{L}g_k z^{-k}]+\sum_{k=1}^{L}\Delta g_k z^{-k}-\sum_{k=1}^{L}k\,g_k\,z^{-k-1}\Delta z \;\approx\; 0\,.$$

Therefore, the first order perturbation of the estimated i-th zero is

$$\Delta z_i \;\approx\; \frac{\sum_{k=1}^{L}\Delta g_k z_i^{-k}}{\sum_{k=1}^{L}k\,g_k z_i^{-k-1}} \;=\; \frac{\mathbf{z}_i^{H}\Delta\mathbf{g}}{\sum_{k=1}^{L}k\,g_k z_i^{-k-1}}\,, \qquad (A13.1)$$

where $\mathbf{z}_i^{H} = [\, z_i^{-1}, z_i^{-2}, \ldots, z_i^{-L}\,]$, $\Delta\mathbf{g} = [\,\Delta g_1, \Delta g_2, \ldots, \Delta g_L\,]^{T}$.

Next we derive the first order perturbation of Equation (4.2.3) with $\hat{M} = M$:

$$\hat{\mathbf{g}} \;=\; -\hat{A}_M^{+}\,\hat{\mathbf{h}}\,, \qquad (4.2.3)$$

$$\begin{aligned}
\therefore \quad \hat{\mathbf{g}} &= -[\,A^{+}+\Delta A_M^{+}\,]\,(\,\mathbf{h}+\Delta\mathbf{h}\,)\\
&\approx -A^{+}\mathbf{h} - A^{+}\Delta\mathbf{h} - \Delta A_M^{+}\mathbf{h}\\
&= \mathbf{g} - A^{+}\Delta\mathbf{h} - \Delta A_M^{+}\mathbf{h}\,.
\end{aligned}$$

$$\therefore \quad \Delta\mathbf{g} \;=\; \hat{\mathbf{g}}-\mathbf{g} \;=\; -A^{+}\Delta\mathbf{h} - \Delta A_M^{+}\mathbf{h}\,. \qquad (A13.2)$$

If $\mathbf{z}_i^{H}$ is the i-th row of Z_R defined by Equation (4.2.2.13), then one can show that the first order perturbation of z_i is given by Equation (4.2.2.3),

$$\Delta z_i = \frac{-1}{|a_i|} \frac{\mathbf{p}_i^H \Delta \boldsymbol{A}' \mathbf{g}'}{\sum_{k=1}^{L} k\, g_k\, z_i^{-k-1}} , \qquad (4.2.2.3)$$

where $\mathbf{p}_i^H$ is the i-th row of the pseudo-inverse of $\boldsymbol{Z}_L$ defined by Equation (4.2.2.11). $\mathbf{g}' = [\ 1, \mathbf{g}^T\ ]$, and $\Delta\boldsymbol{A}'$ is matrix filled with noise components, i.e.,

$$\Delta \boldsymbol{A}' = [\ \Delta\mathbf{h}, \Delta\boldsymbol{A}\ ]$$

$$= \begin{bmatrix} w^*(0) & w^*(1) & \ldots. & w^*(L) \\ w^*(1) & w^*(2) & \ldots. & w^*(L+1) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ w^*(N-L-1) & w^*(N-L) & \ldots & w^*(N-1) \end{bmatrix} , \qquad (4.2.2.6)$$

where the w(k)'s are noise added to the system.

If the additive noise w(k)'s are zero mean white noise, with the real and imaginary part uncorrelated and have variance $\sigma_W{}^2$ for each part. In other words,

$$E\{w(k)\} = 0 , \qquad (4.2.2.22)$$

$$E\{w(k)w(l)\} = 0 , \qquad (4.2.2.23)$$

$$E\{w(k)w^*(l)\} = 2\,\sigma_W{}^2\,\delta_{k,l} , \qquad (4.2.2.24)$$

where $\delta_{k,l}$ is the Kronecker delta function. Then from Equation (4.2.2.3),

$$E[\Delta z_i] = \frac{-1}{|a_i|} \frac{E[\mathbf{p}_i^H \Delta\boldsymbol{A}' \mathbf{g}']}{\sum_{k=1}^{L} k\, g_k\, z_i^{-k-1}} = 0 , \qquad (A13.3)$$

because

$$E[\mathbf{p}_i^H \Delta\boldsymbol{A}'\, \mathbf{g}'] = \mathbf{p}_i^H\, E[\Delta\boldsymbol{A}']\, \mathbf{g}' = 0 . \qquad (A13.4)$$

Now, since

$$
\begin{aligned}
& E\left[\left|\mathbf{p}_i^H \Delta \boldsymbol{A}' \, \mathbf{g}'\right|^2\right] \\
&\quad = E\left[\, \mathbf{p}_i^H \Delta \boldsymbol{A}' \, \mathbf{g}' \mathbf{g}'^H \, \Delta \boldsymbol{A}'^H \mathbf{p}_i \,\right] \\
&\quad = \mathbf{p}_i^H E\left[\, \Delta \boldsymbol{A}' \, \mathbf{g}' \mathbf{g}'^H \, \Delta \boldsymbol{A}'^H \,\right] \mathbf{p}_i \\
&\quad = \mathbf{p}_i^H E \left\{ \begin{bmatrix} \sum_{k=0}^{L} g_k w^*(k) \\ \vdots \\ \sum_{k=0}^{L} g_k w^*(N-L-1+k) \end{bmatrix} \left[\sum_{k=0}^{L} g_k^* w(k), \; \cdots \; , \sum_{k=0}^{L} g_k^* w(N-L-1+k) \right] \right\} \mathbf{p}_i \\
&\quad = \mathbf{p}_i^H \, (2\sigma_W^2 \boldsymbol{R}_g) \, \mathbf{p}_i \; , \qquad \text{(A13.5)}
\end{aligned}
$$

where

$$
(\boldsymbol{R}_g)_{i,j} = (\boldsymbol{R}_g)_{i-j} = (\boldsymbol{R}_g)^*_{j,i} = \begin{cases} \sum_{k=i-j}^{L} g_k g^*_{k-(i-j)} \; ; & 0 \le i-j \le L \\ 0 \; ; & \text{otherwise}. \end{cases} \qquad \text{(4.2.2.27)}
$$

Therefore, the variance of Δz_i is given by

$$
\mathrm{Var}(\Delta z_i) = E[\, |\Delta z_i|^2 \,] - E^2[\, \Delta z_i \,] \qquad \text{(A13.6)}
$$

$$
= \frac{1}{|a_i|^2} \, \frac{E[\,|\mathbf{p}_i^H \Delta \boldsymbol{A}' \, \mathbf{g}'|^2\,]}{\left| \sum_{k=1}^{L} k \, g_k \, z_i^{-k-1} \right|^2} \qquad \text{(A13.7)}
$$

$$
= \frac{2\sigma_W^2}{|a_i|^2} \, \frac{\mathbf{p}_i^H \boldsymbol{R}_g \mathbf{p}_i}{\left| \sum_{k=1}^{L} k \, g_k \, z_i^{-k-1} \right|^2} \qquad \text{(A13.8)}
$$

$$
= \frac{2\sigma_W^2}{|a_i|^2} \, \frac{\mathbf{p}_i^H \boldsymbol{G} \boldsymbol{G}^H \mathbf{p}_i}{\left| \sum_{k=1}^{L} k \, g_k \, z_i^{-k-1} \right|^2} \; , \qquad \text{(A13.9)}
$$

where

$$G = \begin{bmatrix} g_L & \cdot & g_0 & 0 & \ldots & 0 \\ 0 & g_L & \cdot & g_0 & \cdot & \cdot \\ \cdot & & \cdot & & \cdot & 0 \\ 0 & \cdot & 0 & g_L & \cdot & g_0 \end{bmatrix},$$

is an (N-L) x N complex matrix.

Equation (A13.7) is obtained by substituting Equations (A13.3) and (13.5) into Equation (A13.6).

Appendix 14 Proof of property 2 in chapter 4

$\mathrm{Var}(\Delta z_i)$ is invariant to either shifting all frequencies by a constant value or changing all phases by a constant value or both.

Proof :

$$\text{Let} \quad \tilde{f}_i = f_i + C_f \quad \text{and} \quad \tilde{\phi}_i = \phi_i + C_\phi , \tag{A14.1}$$

$$\text{then} \quad \tilde{z}_i = e^{2\pi(-\oe_i + j\tilde{f}_i)} = e^{-2\pi\oe_i + j2\pi(f_i + C_f)} = z_i\, e^{j2\pi C_f} , \tag{A14.2}$$

$$\text{and} \quad 1 + \sum_{k=1}^{L} \tilde{g}_k \tilde{z}^{-k} = 0 \Rightarrow 1 + \sum_{k=1}^{L} \tilde{g}_k z^{-k} e^{-j2\pi k C_f} = 0$$

$$\Rightarrow \tilde{g}_k = g_k\, e^{j2\pi k C_f} . \tag{A14.3}$$

Therefore, from Equation (A14.3) we have

$$\tilde{\mathbf{g}}' = \boldsymbol{D}\, \mathbf{g}' \tag{A14.4}$$

where $\boldsymbol{D} = \mathrm{diag}[1, e^{j2\pi C_f}, e^{j4\pi C_f}, \ldots\ldots, e^{j2\pi L C_f}]$, and $\mathbf{g}' = [1, g_1, g_2, \ldots, g_L]^T$.

Now if

$$\boldsymbol{Z_{B,L}} = \begin{bmatrix} e^{-j\phi_1} & e^{-j\phi_2} & \ldots\ldots & e^{-j\phi_M} \\ e^{s_1^*-j\phi_1} & e^{s_2^*-j\phi_2} & \ldots\ldots & e^{s_M^*-j\phi_M} \\ e^{2s_1^*-j\phi_1} & e^{2s_2^*-j\phi_2} & \ldots\ldots & e^{2s_M^*-j\phi_M} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ e^{(N-L-1)s_1^*-j\phi_1} & e^{(N-L-1)s_2^*-j\phi_2} & \ldots\ldots & e^{(N-L-1)s_M^*-j\phi_M} \end{bmatrix} , \tag{A14.5}$$

where $s_i = 2\pi(-\oe_i + jf_i)$, then

$$\tilde{\boldsymbol{Z}}_{\boldsymbol{B,L}} = \boldsymbol{E_C}\, \boldsymbol{Z_{B,L}} , \tag{A14.6}$$

where $\boldsymbol{E_C} = e^{-jC_\phi}\, [\, \mathrm{diag}(1, e^{-j2\pi C_f}, e^{-j4\pi C_f}, \ldots\ldots, e^{-j2\pi(N-L-1)C_f})\,]$.

Defining $\tilde{\mathbf{p}}_{B,i}^{H}$ as the i-th row of $\tilde{\boldsymbol{Z}}_{B,L}^{+}$, then

$$\begin{aligned} \tilde{\mathbf{p}}_{B,i}^{H} &= \text{i-th row of } (\tilde{\boldsymbol{Z}}_{B,L}^{H} \tilde{\boldsymbol{Z}}_{B,L})^{-1} \tilde{\boldsymbol{Z}}_{B,L}^{H} \\ &= \text{i-th row of } (\boldsymbol{Z}_{B,L}^{H} \boldsymbol{Z}_{B,L})^{-1} \boldsymbol{Z}_{B,L}^{H} \boldsymbol{E}_{C}^{H} e^{jC_{\phi}} \\ &= \text{i-th row of } \mathbf{p}_{B,i}^{H} \boldsymbol{E}_{C}^{H} e^{jC_{\phi}} . \end{aligned} \tag{A14.7}$$

Hence, from Equation (A13.7),

$$\operatorname{Var}(\Delta \tilde{z}_i) = \frac{1}{|a_i|^2} \frac{E[\,|\tilde{\mathbf{p}}_{B,i}^{H} \Delta \boldsymbol{A}' \tilde{\mathbf{g}}'|^2\,]}{\left| \sum_{k=1}^{L} k\, \tilde{g}_k \tilde{z}_i^{-k-1} \right|^2} . \tag{A14.8}$$

From Equations (A14.2) and (A14.3),

$$\begin{aligned} \left| \sum_{k=1}^{L} k\, \tilde{g}_k \tilde{z}_i^{-k-1} \right|^2 &= \left| \sum_{k=1}^{L} k\, g_k e^{j2\pi k C_f} z_i^{-k-1} e^{-j2\pi k C_f - j2\pi C_f} \right|^2 \\ &= \left| \sum_{k=1}^{L} k\, g_k z_i^{-k-1} \right|^2 . \end{aligned} \tag{A14.9}$$

From Equation (A14.7) and Equation (A14.4),

$$\begin{aligned} & E[\,|\tilde{\mathbf{p}}_{B,i}^{H} \Delta \boldsymbol{A}' \tilde{\mathbf{g}}'|^2\,] \\ &\quad = E[\,|\mathbf{p}_{B,i}^{H} \boldsymbol{E}_{C}^{H} e^{jC_{\phi}} \Delta \boldsymbol{A}' \boldsymbol{D}\, \mathbf{g}'|^2\,] \\ &\quad = E[\,(\mathbf{p}_{B,i}^{H} \boldsymbol{E}_{C}^{H} e^{jC_{\phi}} \Delta \boldsymbol{A}' \boldsymbol{D}\, \mathbf{g}')(\mathbf{g}'^{H} \boldsymbol{D}^{H} \Delta \boldsymbol{A}'^{H} e^{-jC_{\phi}} \boldsymbol{E}_{C}\, \mathbf{p}_{B,i})\,] \\ &\quad = \mathbf{p}_{B,i}^{H} \boldsymbol{E}_{C}^{H} E[\,\Delta \boldsymbol{A}' \boldsymbol{D}\, \mathbf{g}' \mathbf{g}'^{H} \boldsymbol{D}^{H} \Delta \boldsymbol{A}'^{H}\,] \boldsymbol{E}_{C}\, \mathbf{p}_{B,i} \\ &\quad = \mathbf{p}_{B,i}^{H} (\, 2\sigma_{W}^{2} \tilde{\boldsymbol{R}}_{g}\,)\, \mathbf{p}_{B,i} , \end{aligned} \tag{A14.10}$$

where the element of the m-th row, n-th column of $\tilde{\boldsymbol{R}}_{g}$ for $0 \le m-n \le L$, is

$$[\tilde{\boldsymbol{R}}_{g}]_{m,n} = \sum_{k=i-j}^{L} \tilde{g}_k \tilde{g}_{k-(m-n)}^{*}$$

$$= \sum_{k=i-j}^{L} g_k \, g^{*}_{k-(m-n)} \, e^{j2\pi(m-n)C_f} .$$

Since,

$$[\boldsymbol{E}_C^{H} \tilde{\boldsymbol{R}}_g \boldsymbol{E}_C]_{m,n} = \sum_{k=i-j}^{L} g_k \, g^{*}_{k-(m-n)} \, e^{j2\pi(m-n)C_f} \, e^{-j2\pi(m-1)C_f} \, e^{j2\pi(n-1)C_f}$$

$$= \sum_{k=i-j}^{L} g_k \, g^{*}_{k-(m-n)}$$

$$= [\boldsymbol{R}_g]_{m,n} ,$$

therefore, from Equation (A14.10),

$$E[\,|\tilde{p}^{H}_{B,i} \, \Delta \boldsymbol{A}' \, \tilde{g}' |^2\,] = p^{H}_{B,i} \, (\, 2\sigma^2_W \boldsymbol{R}_g \,) \, p_{B,i} \, . \qquad \text{(A14.11)}$$

Combining Equation (A14.8), Equation (A14.9) and Equation (A14.11), we conclude that

$$\mathrm{Var}\,(\Delta\tilde{z}_i\,) = \mathrm{Var}\,(\Delta z_i\,) \, .$$

Appendix 15 Derivation of Cramer-Rao (CR) bounds for the parameters of M multiple exponentially damped signals in white Gaussian noise

Let the data samples $y(t_n)$ be given by the formula

$$y(t_n) \;=\; \sum_{k=1}^{M} A_k\, e^{j\phi_k}\, e^{(-\alpha_k + j\omega_k)t_n} + w(t_n)\,, \quad 0 \le n \le N-1, \qquad \text{(A15.1)}$$

where $w(t_n)$ is a white Gaussian sequence. We will assume a constant sampling rate of $1/\Delta t$ [Hz] with the first sample taken at $t = t_0$. Thus

$$t_n \;=\; t_0 + n\Delta t \;=\; (n_0 + n)\,\Delta t \;\; \text{[sec]}.$$

Each real and imaginary part of $w(t_n)$ has a zero mean and variance $\sigma^2 = \frac{{\sigma_w}^2}{2}$. If we write

$$y(t_n) \;=\; x_n + j\, z_n\,,$$

then the joint probability density function (PDF) of the elements of the sample vector **y** when the unknown parameter vector is $\boldsymbol{\alpha}$ is given by

$$\begin{aligned} f(\mathbf{y};\boldsymbol{\alpha}) &= \left(\frac{1}{2\pi\sigma^2}\right)^N \exp\left[\frac{-1}{2\sigma^2} \sum_{n=0}^{N-1} \left| y(t_n) - \sum_{k=1}^{M} A_k\, e^{j\phi_k}\, e^{(-\alpha_k + j\omega_k)\, t_n} \right|^2 \right] \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^N \exp\left[\frac{-1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - \mu_n)^2 + (z_n - \gamma_n)^2 \right] \qquad \text{(A15.2)} \end{aligned}$$

where, if A_k, ϕ_k, α_k, and ω_k, for $k = 1,2, \ldots , M$, are all unknown,

$$\boldsymbol{\alpha} \;=\; [A_1, \phi_1, \alpha_1, \omega_1, A_2, \phi_2, \alpha_2, \omega_2, \ldots\ldots\ldots\ldots, A_M, \phi_M, \alpha_M, \omega_M]^T,$$

$$\mu_n \;=\; \sum_{k=1}^{M} A_k\, e^{-\alpha_k t_n} \cos(\omega_k t_n + \phi_k)\,,$$

$$\gamma_n \;=\; \sum_{k=1}^{M} A_k\, e^{-\alpha_k t_n} \sin(\omega_k t_n + \phi_k)\;.$$

The unbiased CR bound are the diagonal elements of the inverse of the Fisher information matrix $\mathfrak{J}$, whose elements are given by

$$[\mathfrak{I}]_{ij} = -E\left[\frac{\partial^2 \ln f(\mathbf{y};\boldsymbol{\alpha})}{\partial\alpha_i\,\partial\alpha_j}\right], \qquad i, j = 1, 2, \ldots\ldots, 4M \tag{A15.3}$$

where the expectation E[ ] is with respect to the sample vector **y**. The variance of the estimated parameters is then given by

$$\operatorname{var}\{\hat{\alpha}_i\} \geq [\mathfrak{I}^{-1}]_{ii} \tag{A15.4}$$

where $\hat{\alpha}_i$ is the estimator of α_i and $[\mathfrak{I}^{-1}]_{ii}$ is the i-th diagonal element of $\mathfrak{I}^{-1}$.

If $f(\mathbf{y};\boldsymbol{\alpha})$ is given by Equation (A15.2), then one can show that the elements of $\mathfrak{I}$ can be written as

$$[\mathfrak{I}]_{ij} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left[\frac{\partial\mu_n}{\partial\alpha_i}\frac{\partial\mu_n}{\partial\alpha_j} + \frac{\partial\gamma_n}{\partial\alpha_i}\frac{\partial\gamma_n}{\partial\alpha_j}\right]. \tag{A15.5}$$

The first-order partial derivatives are given by

$$\begin{aligned}
\frac{\partial\mu_n}{\partial A_i} &= e^{-\alpha_i t_n}\cos(\omega_i t_n + \phi_i)\\
\frac{\partial\mu_n}{\partial\phi_i} &= -A_i e^{-\alpha_i t_n}\sin(\omega_i t_n + \phi_i)\\
\frac{\partial\mu_n}{\partial\alpha_i} &= -t_n A_i e^{-\alpha_i t_n}\cos(\omega_i t_n + \phi_i)\\
\frac{\partial\mu_n}{\partial\omega_i} &= -t_n A_i e^{-\alpha_i t_n}\sin(\omega_i t_n + \phi_i)\\
\frac{\partial\gamma_n}{\partial A_i} &= e^{-\alpha_i t_n}\sin(\omega_i t_n + \phi_i)\\
\frac{\partial\gamma_n}{\partial\phi_i} &= A_i e^{-\alpha_i t_n}\cos(\omega_i t_n + \phi_i)\\
\frac{\partial\gamma_n}{\partial\alpha_i} &= -t_n A_i e^{-\alpha_i t_n}\sin(\omega_i t_n + \phi_i)\\
\frac{\partial\gamma_n}{\partial\omega_i} &= t_n A_i e^{-\alpha_i t_n}\cos(\omega_i t_n + \phi_i).
\end{aligned} \tag{A15.6}$$

From this, one can show that the Fisher information matrix is as follows:

$$\mathfrak{I} \quad = \quad \frac{1}{\sigma^2}\begin{bmatrix} \mathfrak{I}_{11} & \mathfrak{I}_{12} & \cdots\cdots & \mathfrak{I}_{1M} \\ \mathfrak{I}_{21} & \mathfrak{I}_{22} & \cdots\cdots & \mathfrak{I}_{2M} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathfrak{I}_{M1} & \mathfrak{I}_{M2} & \cdots\cdots & \mathfrak{I}_{MM} \end{bmatrix}, \qquad \text{(A15.7)}$$

where $\mathfrak{I}_{ij}$ is the following 4 x 4 matrix

$$\mathfrak{I}_{ij} = \sum_{n=0}^{N-1} \begin{bmatrix} \psi_{ijn} & A_j\xi_{ijn} & -A_j t_n \psi_{ijn} & A_j t_n \xi_{ijn} \\ -A_i\xi_{ijn} & A_iA_j\psi_{ijn} & A_iA_j t_n \xi_{ijn} & A_iA_j t_n \psi_{ijn} \\ -A_i t_n \psi_{ijn} & -A_iA_j t_n \xi_{ijn} & A_iA_j t_n^2 \psi_{ijn} & -A_iA_j t_n^2 \xi_{ijn} \\ -A_i t_n \xi_{ijn} & A_iA_j t_n \psi_{ijn} & A_iA_j t_n^2 \xi_{ijn} & A_iA_j t_n^2 \psi_{ijn} \end{bmatrix}$$

and

$$\psi_{ijn} \quad = \quad e^{-(\alpha_i+\alpha_j)t_n} \cos\Delta_n[i,j] \ ,$$

$$\xi_{ijn} \quad = \quad e^{-(\alpha_i+\alpha_j)t_n} \sin\Delta_n[i,j] \ ,$$

$$\Delta_n[i,j] = \quad (\omega_i-\omega_j)t_n + (\phi_i-\phi_j) \quad .$$

The matrix $\mathfrak{I}$ can then be inverted to compute the CR bounds of the estimated parameters $\hat{A}_k$, $\hat{\phi}_k$, $\hat{\alpha}_k$ and $\hat{\omega}_k$, for k = 1,2, ... , M. The results are as follows :

$$\mathrm{var}(\hat{A}_k) \quad \geq \quad \sigma^2\,[\mathfrak{I}^{-1}]_{(4k-3)(4k-3)} \ , \qquad \text{(A15.8)}$$

$$\mathrm{var}(\hat{\phi}_k) \quad \geq \quad \sigma^2\,[\mathfrak{I}^{-1}]_{(4k-2)(4k-2)} \quad [\mathrm{rad}^2], \qquad \text{(A15.9)}$$

$$\mathrm{var}(\hat{\alpha}_k) \quad \geq \quad \sigma^2\,[\mathfrak{I}^{-1}]_{(4k-1)(4k-1)} \quad [(\mathrm{rad/sec})^2]$$

$$= \quad \frac{\sigma^2}{(2\pi)^2}\,[\mathfrak{I}^{-1}]_{(4k-1)(4k-1)} \quad [\mathrm{Hz}^2] \ , \qquad \text{(A15.10)}$$

$$\mathrm{var}(\hat{\omega}_k) \quad \geq \quad \sigma^2\,[\mathfrak{I}^{-1}]_{(4k)(4k)} \quad [(\mathrm{rad/sec})^2]$$

$$= \quad \frac{\sigma^2}{(2\pi)^2(\mathrm{SF}+\mathrm{O1})^2}\,[\mathfrak{I}^{-1}]_{(4k)(4k)} \quad [\mathrm{ppm}^2] \ . \qquad \text{(A15.11)}$$

Appendix 16 Derivation of the magnitude square area of an exponentially decaying signal peak

Let the time domain exponentially decaying signal be

$$y(t) \quad = \quad A_0\, e^{\,j\phi_0}\, e^{\,(-2\pi\alpha_0 + j\omega_0)t}, \qquad \text{(A16.1)}$$

for $t \geq 0$. Denote the Fourier transformation of $y(t)$ by $Y(\omega)$, then

$$\begin{aligned}
Y(\omega) \quad &\equiv \quad \int_{-\infty}^{\infty} y(t)\, e^{-j\omega t}\, dt \\
&= \quad \int_{0}^{\infty} A_0 e^{\phi_0}\, e^{(-2\pi\alpha_0 + j\omega_0)t}\, e^{-j\omega t}\, dt \\
&= \quad A_0 e^{j\phi_0} \left[\frac{e^{-2\pi\alpha_0 t + j(\omega_0 - \omega)t}}{-2\pi\alpha_0 + j(\omega_0 - \omega)} \right]_{t=0}^{\infty} \\
&= \quad \frac{A_0 e^{j\phi_0}}{2\pi\alpha_0 - j(\omega_0 - \omega)}
\end{aligned}$$

$$= \frac{2\pi A_0\alpha_0\cos\phi_0 - A_0(\omega_0-\omega)\sin\phi_0}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2} + j\,\frac{2\pi A_0\alpha_0\sin\phi_0 + A_0(\omega_0-\omega)\cos\phi_0}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2}. \qquad \text{(A16.2)}$$

$\therefore$ From Equation (A16.2), we have

$$|Y(\omega)|^2 \quad = \quad Y^*(\omega)Y(\omega) \quad = \quad \frac{A_0^2}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2}.$$

The area under $|Y(\omega)|^2$ is then given by

$$\int_{-\infty}^{\infty} |Y(\omega)|^2\, d\omega \quad = \quad \int_{-\infty}^{\infty} \frac{A_0^2\, d\omega}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2} \quad = \quad \frac{A_0^2}{2\alpha_0}. \qquad \text{(A16.3)}$$

Appendix 17 Derivation of the area under the real part of the spectrum for an exponential decaying signal

Let the time domain exponentially decaying signal be defined by Equation (A16.1). Then from Equation (A16.2), the real part of the frequency spectrum is

$$\mathrm{Re}\,(Y(\omega)) = \frac{2\pi A_0\alpha_0\cos\phi_0 - A_0(\omega_0 - \omega)\sin\phi_0}{(2\pi\alpha_0)^2 + (\omega_0 - \omega)^2} . \qquad \text{(A17.1)}$$

We will consider two cases: first the spectrum has no phase error, and secondly, the spectrum has phase error.

(a) Re(Y(ω)) has no phase error , i.e., $\phi_0 = 0$, then the area under Re(Y(ω)) is given by

$$\int_{-\infty}^{\infty} \mathrm{Re}(Y(\omega))\, d\omega = \int_{-\infty}^{\infty} \frac{2\pi\, A_0\, \alpha_0\, d\omega}{(2\pi\alpha_0)^2 + (\omega_0\text{-}\omega)^2} = A_0\, \pi\, . \qquad \text{(A17.2)}$$

(b) Re(Y(ω)) has phase error , i.e., $\phi_0 \neq 0$, then the area under Re(Y(ω)) is given by

$$\int_{-\infty}^{\infty} \mathrm{Re}(Y(\omega))\, d\omega = \int_{-\infty}^{\infty} \frac{2\pi A_0\alpha_0\cos\phi_0 - A_0(\omega_0\text{-}\omega)\sin\phi_0}{(2\pi\alpha_0)^2 + (\omega_0\text{-}\omega)^2}\, d\omega\, . \qquad \text{(A17.3)}$$

Since, from Equation (A17.2),

$$\int_{-\infty}^{\infty} \frac{d\omega}{(2\pi\alpha_0)^2 + (\omega_0\text{-}\omega)^2} = \frac{\pi}{2\pi\alpha_0}\, , \qquad \text{(A17.4)}$$

and

$$\lim_{B\to\infty} \int_{-B}^{B} \frac{x\, dx}{(2\pi\alpha_0)^2 + x^2} = \lim_{B\to\infty} \ln\left((2\pi\alpha_0)^2 + x^2\right)\Big|_{x=-B}^{B} = 0\, . \qquad \text{(A17.5)}$$

Therefore, substituting Equations (A17.4) and (A17.5) into Equation (A17.3) results in

$$\int_{-\infty}^{\infty} \mathrm{Re}(Y(\omega))\, d\omega = A_0 \cos\phi_0\, \pi . \tag{A17.6}$$

Appendix 18 Linewidth of an exponentially damped signal

Let the time domain exponentially decaying signal be defined by Equation (A16.1), then its Fourier transformation is given by Equation (A16.2), i.e.,

$$Y(\omega) = \frac{2\pi A_0\alpha_0\cos\phi_0 - A_0(\omega_0-\omega)\sin\phi_0}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2} + j\,\frac{2\pi A_0\alpha_0\sin\phi_0 + A_0(\omega_0-\omega)\cos\phi_0}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2} \,. \tag{A18.1}$$

$$\therefore \quad |\,Y(\omega)\,| = \sqrt{\frac{A_0^2}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2}} \,, \tag{A18.2}$$

and so, $|Y(\omega)|$ at ω_0 is

$$|Y(\omega_0)| = \frac{A_0}{2\pi\alpha_0} \,. \tag{A18.3}$$

To find the frequencies where the magnitude is half-height of that from Equation (A18.3), solve the following equation for ω:

$$\frac{1}{2}\,\frac{A_0}{2\pi\alpha_0} = \sqrt{\frac{A_0^2}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2}} \,,$$

which results in

$$\omega = \omega_0 \pm 2\pi\sqrt{3}\,\alpha_0 \,. \tag{A18.4}$$

Therefore, the full-width at half-height for magnitude spectrum is

$$\Delta\omega = 2\sqrt{3}\,2\pi\,\alpha_0 \,. \tag{A18.5}$$

Now consider the real part of the spectrum with $\phi_0 = 0$. We have from Equation (A18.1) that

$$\mathrm{Re}\,(Y(\omega)) = \frac{2\pi A_0\alpha_0}{(2\pi\alpha_0)^2 + (\omega_0-\omega)^2} \,, \tag{A18.6}$$

and so,

$$\mathrm{Re}\,(Y(\omega_0)) = \frac{A_0}{2\pi\alpha_0} \,. \tag{A18.7}$$

To find the frequencies where the real part of the spectrum is half-height of that from Equation (A18.6), solve the following equation for ω:

$$\frac{1}{2}\frac{A_0}{2\pi\alpha_0} = \frac{2\pi A_0 \alpha_0}{(2\pi\alpha_0)^2 + (\omega_0 - \omega)^2} .$$

The result is as follows :

$$\omega = \omega_0 \pm 2\pi\alpha_0 . \tag{A18.8}$$

Hence, from Equation (A18.8), the full-width at half-height for the real part of the spectrum with no phase error is

$$\Delta\omega^* = 4\pi\,\alpha_0 . \tag{A18.9}$$

Appendix 19 Derivation of time domain signal-to-noise ratio for an exponentially decaying signal embedded in complex white Gaussian noise

Let us denote the time domain exponentially decaying signal by

$$y(n) \quad = \quad A_0 e^{j\phi_0} e^{(-2\pi\alpha_0 + j\omega_0)n\Delta t}, \qquad n = 0, \ldots, N-1, \qquad \text{(A19.1)}$$

where $\alpha_0 > 0$. We further assume that the noise sequence w(n) is white complex Gaussian noise with uncorrelated real and imaginary parts each having a variance of σ_w^2, i.e.,

$$E\{w(k)\} \quad = \quad 0, \qquad \text{(A19.2)}$$

$$E\{w(k)w(l)\} \quad = \quad 0, \qquad \text{(A19.3)}$$

$$E\{w(k)w^*(l)\} \quad = \quad 2\,\sigma_W^2\,\delta_{k,l}, \qquad \text{(A19.4)}$$

where $\delta_{k,l}$ is the Kronecker delta function and $E\{w(k)\}$ is the ensemble average of w(k)'s. Define the time domain signal-to-noise (SNR) to be

$$\begin{aligned}(SNR)_t \quad &\equiv \quad \frac{\text{average power of signal y}}{\text{average power of noise w}} \\ &= \quad \frac{E\{P_y\}}{E\{P_w\}}. \qquad \text{(A19.5)}\end{aligned}$$

Now,

$$\begin{aligned}E\{P_y\} \quad &= \quad E\{\frac{1}{N\Delta t}\sum_{n=0}^{N-1} |y(n)|^2\} \\ &= \quad \frac{1}{N\Delta t}\sum_{n=0}^{N-1} |A_0|^2 e^{-4\pi\alpha_0 n\Delta t} \\ &= \quad \frac{|A_0|^2}{N\Delta t}\left[\frac{1-e^{-4\pi\alpha_0 N\Delta t}}{1-e^{-4\pi\alpha_0\Delta t}}\right], \qquad \text{since } \alpha_0 \neq 0, \\ &\approx \quad \frac{|A_0|^2}{N\Delta t}\,\frac{1}{1-[1-4\pi\alpha_0\Delta t+\ldots]}, \qquad \text{for } -4\pi\,\alpha_0\,\Delta t << 0 \\ &= \quad \frac{|A_0|^2}{N\Delta t}\,\frac{1}{4\pi\alpha_0\Delta t}. \qquad \text{(A19.6)}\end{aligned}$$

Next,

$$
\begin{aligned}
E\{P_w\} &= E\left\{ \frac{1}{N\Delta t} \sum_{n=0}^{N-1} |w(n)|^2 \right\} \\
&= \frac{1}{N\Delta t} \sum_{n=0}^{N-1} E\{|w(n)|^2\} \\
&= \frac{1}{N\Delta t} \, 2\sigma_w^2 \, N \\
&= \frac{2\sigma_w^2}{\Delta t} \, . \qquad \text{(A19.7)}
\end{aligned}
$$

Therefore, substituting both Equations (A19.6) and (A19.7) into Equation (A19.5) results in

$$
(SNR)_t = \frac{1}{2N} \, \frac{1}{2\pi\alpha_0\Delta t} \, \frac{|A_0|^2}{2\sigma_w^2} \, . \qquad \text{(A19.8)}
$$

Appendix 20 Statistical relationship of zero mean white Gaussian noise in time domain and frequency domain

In this appendix, we will show the statistical relationship of sequences of zero mean complex white Gaussian noise in time domain and frequency domain. By statistical relationship, we mean the average and standard deviation.

Let us denote w(n), for n = 0, 1, . . ., N-1, to be a sequence of white complex Gaussian noise with zero mean, with uncorrelated real and imaginary parts each having a variance of σ_W^2, i.e.,

$$E\{w(k)\} = 0\ , \tag{A20.1}$$

$$E\{w(k)w(l)\} = 0\ , \tag{A20.2}$$

$$E\{w(k)w^*(l)\} = 2\,\sigma_W{}^2\,\delta_{k,l}\ , \tag{A20.3}$$

where $\delta_{k,l}$ is the Kronecker delta function and $E\{w(k)\}$ is the ensemble average of w(k)'s. From Equation (A4.2), the DFT of w(n) is

$$W(mF) = \Delta t \sum_{n=0}^{N-1} w(n)\, e^{-j2\pi nm/N}\ , \tag{A20.4}$$

where $F = \frac{1}{N\Delta t}$ and Δt is the sampling interval in the time domain.

If we take the ensemble average of Equation (A20.4) and use Equation (A20.1), we get

$$E\{W(mF)\} = \Delta t \sum_{n=0}^{N-1} E\{w(n)\}\, e^{-j2\pi nm/N} = 0. \tag{A20.5}$$

Next, the variance and covariance of the estimator in Equation (A20.4) must be derived. The covariance is defined as

$$\mathrm{cov}\,\{W(mF),W(kF)\} = E\{W^*(mF)W(kF)\} - E\{W(mF)\}\,E\{W(kF)\}. \tag{A20.6}$$

From Equation (A20.5) the last term in Equation (A20.6) is zero, so we only need to compute $E\{W^*(mF)W(kF)\}$. If Equation (A20.4) is used, we get

$$E\{W^*(mF)W(kF)\} = (\Delta t)^2 \sum_{n=0}^{N-1} \sum_{l=0}^{N-1} E\{w^*(n)w(l)\}\, e^{-j2\pi nm/N}\, e^{-j2\pi kl/N}. \tag{A20.7}$$

Substituting Equation (A20.3) into the above equation we have

$$E\{W^*(mF)W(kF)\} = (\Delta t)^2 \sum_{n=0}^{N-1} 2\sigma_W^2\, e^{-j2\pi n(m-k)/N}. \tag{A20.8}$$

Therefore, from Equation (A20.6), (A20.5) and (A20.8), we find

$$\mathrm{cov}\,\{W(mF), W(nF)\} = (\Delta t)^2\, 2\sigma_W^2 \sum_{n=0}^{N-1} e^{-j2\pi n(m-k)/N}. \tag{A20.9}$$

Let us now use Equation (A20.9) to study the variance of the estimator in Equation (A20.4). If m = k, Equation (A20.9) becomes

$$\mathrm{var}\,\{W(mF)\} = (\Delta t)^2\, 2\sigma_W^2\, N, \tag{A20.10}$$

which is equal to 2 times the variance of w(n) in the frequency domain, $2(\sigma_W)_f^2$, because the transformed noise sequence is still white complex Gaussian with zero mean, with uncorrelated real and imaginary parts each having a variance of $(\sigma_W)_f^2$. That is,

$$\mathrm{var}\,\{W(mF)\} \equiv 2(\sigma_W)_f^2 = (\Delta t)^2\, 2\sigma_W^2\, N. \tag{A20.11}$$

Therefore, the standard deviation of the noise in the real part of the estimator in Equation (A20.4) is given by

$$(\sigma_W)_f = \Delta t\, \sigma_W \sqrt{N}\,. \tag{A20.12}$$

Appendix 21 Levenberg-Marquardt method of nonlinear least squares problem

Assume that the model to be fitted is

$$y = y(x; \mathbf{a}) \tag{A21.1}$$

and the χ^2 merit function is

$$\chi^2(\mathbf{a}) = \sum_{i=1}^{N} [\, y_i - y(x_i; \mathbf{a}) \,]^2 . \tag{A21.2}$$

The gradient of the χ^2 merit function with respect to the parameters a has components

$$\frac{\partial \chi^2(\mathbf{a})}{\partial a_k} = -2 \sum_{i=1}^{N} [\, y_i - y(x_i; \mathbf{a}) \,] \frac{\partial y(x_i; \mathbf{a})}{\partial a_k}, \quad \text{for } k = 1, 2, \ldots, M. \tag{A21.3}$$

The second partial derivative of χ^2 with respect to **a** can be approximated by

$$\frac{\partial^2 \chi^2(\mathbf{a})}{\partial a_k \, \partial a_l} \approx 2 \sum_{i=1}^{N} \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \frac{\partial y(x_i; \mathbf{a})}{\partial a_l} \right]. \tag{A21.4}$$

If the current parameters $\mathbf{a}_i$ are good enough so that the χ^2 merit function is sufficiently close to the minimum, then the χ^2 function may be well approximated by its second order Taylor series expansion around the point $\mathbf{a}_i$:

$$\chi_i^2(\mathbf{a}) = \chi^2(\mathbf{a}_i) + \mathbf{q}_i^T(\mathbf{a} - \mathbf{a}_i) + \frac{1}{2}(\mathbf{a} - \mathbf{a}_i)^T \boldsymbol{H}_i(\mathbf{a} - \mathbf{a}_i), \tag{A21.5}$$

where $\boldsymbol{H}_i$ is the Hessian matrix of χ^2 evaluated at $\mathbf{a} = \mathbf{a}_i$, whose k-th row and l-th column element is

$$[\boldsymbol{H}_i]_{kl} \equiv \left| \frac{\partial^2 \chi^2(\mathbf{a})}{\partial a_k \, \partial a_l} \right|_{\mathbf{a} = \mathbf{a}_i}, \tag{A21.6}$$

and $\mathbf{q}_i$ is the gradient of χ^2 evaluated at $\mathbf{a} = \mathbf{a}_i$, whose k-th element is

$$[\mathbf{q}_i]_k \equiv \left| \frac{\partial \chi^2(\mathbf{a})}{\partial a_k} \right|_{\mathbf{a} = \mathbf{a}_i}. \tag{A21.7}$$

If the approximation is a good one, then we know how to jump from the current trial parameters $\mathbf{a}_i$ to the minimizing ones $\mathbf{a}_{min}$ in a single leap as in the following. We equate to zero the gradient of $\chi_i^2(\mathbf{a})$

$$\frac{\partial \chi_i^2(\mathbf{a})}{\partial a_k} = \mathbf{q}_i + \boldsymbol{H}_i(\mathbf{a} - \mathbf{a}_i) = 0 , \tag{A21.8}$$

which, if $\boldsymbol{H}_i$ is nonsingular, has the solution

$$\mathbf{a}_{min} = \mathbf{a}_{i+1} = \mathbf{a}_i - \boldsymbol{H}_i^{-1}\mathbf{q}_i . \tag{A21.9}$$

On the other hand, Equation (A21.5) might be a poor local approximation to the shape of the function that we are trying to minimize at $\mathbf{a}_i$. In that case, we can take a step down the gradient, as follows:

$$\mathbf{a}_{i+1} = \mathbf{a}_i - \text{constant} \times \mathbf{q}_i , \tag{A21.10}$$

where the constant is small enough not to exhaust the downhill direction.

Combining Equation (A21.9) with Equation (A21.4), we have

$$\sum_{j=1}^{M} [\boldsymbol{H}_i]_{kj}\, \delta\mathbf{a}_j = [\mathbf{q}_i]_k , \tag{A21.11}$$

where $\delta\mathbf{a}_j = [\mathbf{a}_{i+1}]_j - [\mathbf{a}_i]_j$.

Equation (A21.10) can also be translated to

$$\delta\mathbf{a}_j = \text{constant} \times [\mathbf{q}_i]_j . \tag{A21.12}$$

Marquardt (1963) put forth an elegant method, related to an earlier suggestion of Levenberg, for varying smoothly between the extremes of the inverse Hessian method (Equation (A21.11)) and the steepest descent method (Equation (A21.12)) as follows: the constant in Equation (A21.12) is set to be proportional to the j-diagonal element of the Hessian matrix evaluated at $\mathbf{a} = \mathbf{a}_i$

$$\delta\mathbf{a}_j = \frac{1}{\lambda[\boldsymbol{H}_i]_{jj}} [\mathbf{q}_i]_j , \tag{A21.13}$$

where λ is some non-dimensional factor which has the possibility of being set far greater than 1 to cut down the step.

If we define a new matrix $\boldsymbol{H}'$ by the following prescription:

$$[\boldsymbol{H}_i']_{jj} = [\boldsymbol{H}_i]_{jj}\,(1+\lambda)$$
$$[\boldsymbol{H}_i']_{jk} = [\boldsymbol{H}_i]_{jk} \quad , \quad \text{for } (j \neq k)\,. \qquad \text{(A21.14)}$$

Then Equations (A21.11) and (A21.13) can be replaced by

$$\sum_{j=1}^{M} [\boldsymbol{H}_i']_{kj}\,\delta\mathbf{a}_j = [\mathbf{q}_i]_k\,. \qquad \text{(A21.15)}$$

When λ is very large, the matrix $\boldsymbol{H}'$ is forced into being diagonally dominant, so Equation (A21.15) goes over to be identical to Equation (A21.13). On the other hand, as λ approaches zero, Equation (A21.15) goes over to Equation (A21.11).

Given an initial guess for the set of fitted parameters **a**, the recommended Marquardt recipe is as follows:

1. Compute $\chi^2(\mathbf{a})$.
2. Pick a modest value for λ, say $\lambda = 0.001$.
3. Solve Equation (A21.15) for $\delta\mathbf{a}$ and evaluate $\chi^2(\mathbf{a} + \delta\mathbf{a})$.
4. If $\chi^2(\mathbf{a} + \delta\mathbf{a}) \geq \chi^2(\mathbf{a})$, increase λ by a factor of 10 (or any other substantial factor) and go back to step 3.
5. If $\chi^2(\mathbf{a} + \delta\mathbf{a}) < \chi^2(\mathbf{a})$, decrease λ by a factor of 10, update the trial solution **a** to $\mathbf{a} + \delta\mathbf{a}$, and go back to step 3.

The iteration may be stopped on the first or second occasion that χ^2 decreases by a negligible amount.